FIRST DRAFT
(Not for circulation)

TEACHERS' TRAINING MATERIALS IN PLUS TWO MATHEMATICS

DEPARTMENT OF EDUCATION IN SCIENCE AND MATHEMATICS
National Council of Educational Research and Training
NIE Campus : Sri Aurobindo Marg
New Delhi-110 016

FOREWORD

This is perhaps for the first time that the Department of Education in Science and Mathematics, NCERT, has developed teachers' training materials mainly in the thrust areas of plus two mathematics for the benefit of the teacher-educators. After the publication of the textbooks of mathematics on the basis of the new curriculum prepared on the philosophy of the National Policy on Education, 1986, it has been felt that extensive training should be given to the teachers most of whom are not familiar with the new topics and concepts introduced in plus two mathematics. With this objective in view, the training materials have been developed particularly in the new topics in a workshop held in Calcutta in September 1990, in order to facilitate the teacher-educators.

This is only the first draft of the training materials which were developed in the workshop on the basis of a sample paper on a particular topic, prepared by Prof. S.C. Das of the Department, who is the Coordinator of the programme. These draft materials will be exposed to the teacher-educators who will be trained later by the faculty members. After reviewing the materials thoroughly on the basis of suggestions from the experts, the materials will be revised and printed.

I am thankful to all the authors who developed different topics on this draft and my colleagues, Prof. S.C. Das and Dr. Hukum Singh who took active part in the discussion of various topics in the workshop.

Suggestions from the readers for further improvement of this draft will be highly appreciated.

(B. GANGULY)
Head, DESM and Dean
NCERT

New Delhi
December, 1990.

I N D E X

A SAMPLE PAPER

ON

THEORY OF MAXIMUM AND MINIMUM

Prepared by

Professor S. C. Das

(Coordinator of the Programme)

DEPARTMENT OF EDUCATION IN SCIENCE & MATHEMATICS
(NATIONAL COUNCIL OF EDUCATIONAL RESEARCH AND TRAINING)
N I E CAMPUS - SRI AUROBINDO MARG
NEW DELHI - 110016

THEORY OF MAXIMUM AND MINIMUM

1. Motivation of the topic:

We are all familiar with the literal meaning of the terms "maximum" and "minimum". We say that the maximum mark in Mathematics obtained by the students of a school X is 100. We also say that the "minimum" teperature of Srinagar in winter goes below $0^{\circ}C$. It will be explained later on that we distinguish different kinds of maxima and minima in mathematics. Assuming for the time being the literal meaning of "maximum" and "minimum" as used in the above two examples, we may come across with some daily life problems involving "maximum" or "minimum", for the solution of which the elementary mathematics taught upto class XI may not be adequate. We have to take help of the theory of maxima and minima which involves calculus to solve such problems.

For example, suppose that we want to make a rectangular box of maximum capacity out of a given rectangular sheet of tin whose length and breadth are 'a' m and 'b' m respectively by cutting off four equal squares from its four corners and by folding the remaining portions which were attached with these four squares. The question is : What will be the side of each of such square so that the volume is maximum ? Without going into the details of mathematics, if we apply our common sense we may be tempted to say that if the wastage of the tin sheet is minimum, then the capacity of the box may be maximum and so the shorter will be sidesof the squares cut off, the bigger will be the volume of the box.

But after a little thought, it will be clear that this conclusion may not be true, for the shorter is the side of each squares cut off, the shorter will be the depth of the box although the area of the base of the box will be larger. On the other hand, if the sides of the squares cut off are larger and larger, the height of the box will be larger and larger but the area of the base of the box will be smaller and smaller and hence in any of these two cases the volume of the box may not be maximum.

If we now want to solve this problem by using elementary mathematics (pre-calculus), we may perhaps fail to solve it. It is only through the theory of maximum and minimum in Calculus that we can solve such problems.

We thus observe that there are many daily-life problems where the theory of maximum and minimum has to be applied for their solutions.

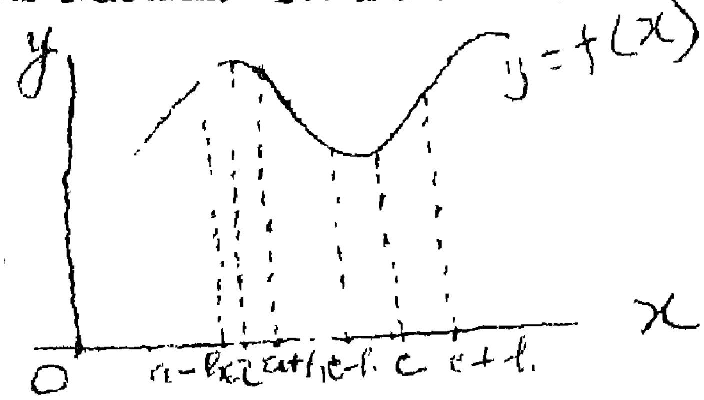
2. Brief outline of the content :

In order to find the local maximum or local minimum value of a function f (x), we find the derivative of f (x), if it exists. We then apply the necessary condition for local maximum and local minimum i.e. we equate $f'(x)$ to zero for local maximum or local minimum and solve for x. Let x_1, x_2 be the solution of this equation. Then $x = x_1$ is the point of maximum if $f''(x_1) < 0$ or minimum if $f''(x_1) >$
Similarly, we test the other point $x = x_2$ for maximum or minimum. The maximum or minimum value of f(x) will thus be $f(x_1)$ or $f(x_2)$.

3. Explanation of technical/mathematical terms not properly explained in the text books

By "maximum" and "minimum" in calculus, we generally mean "local maximum" and "local minimum" which are defined as follows:

Let $y = f(x)$ be a continuous function. Let $x = a$ be a point on the graph such that $f(a) >$ both $f(a-h)$ and $f(a+h)$ for sufficiently small +ve values of h. Then $x = a$ is called a point of local maximum and $f(a)$ is a local maximum value of the function.



Let $x = c$ be a point on the graph such that $f(c) <$ both $f(c-h)$ and $f(c+h)$ for sufficiently small +ve values of h. Then $x = c$ is a point of local minimum and $f(c)$ is called a local minimum value of the function. The points of local maximum and local minimum are also called the points of extremum and the maximum or minimum value of a function is called the extreme value of the function.

There is another type of maximum and minimum called "absolute maximum" and "absolute minimum" of a function by some authors. These are also called "maximum and minimum values of the function in a closed interval" or "the greatest and the least values of the function".

The greatest and the least values of the function can be determined only when the given interval is closed. The greatest value of the function in a closed interval is the greatest ordinate that can be drawn within the interval and the least value of the function is the least value of the ordinate that can be drawn within the interval.

We observe that the greatest and the least values of a function can occur either at the end points of the interval or at those points where there exist local maxima or local minima.

Hence, for the determination of the greatest and the least values of a function f (x) in a given closed interval $a \leq x \leq b$, we first find the points, say $x = x_1$, x_2 etc. $(a \leq x_1, x_2 \leq b)$ by solving $f'(x) = 0$, where the function may have local maxima or local minima. We then determine the ordinates $f(a)$, $f(b)$, $f(x_1)$, $f(x_2)$, etc. Then the greatest and the least of these ordinates will be the greatest and the least values of the function.

The differences between maximum (local)/minimum (local) and the greatest/the least values of a function may be tabulated as follows:

	Maximum/Minimum values (local)		Greatest/Least Values
1.	There may be many maxima & many minima of a function in a given interval.	1.	There may be only one greatest value and only one least value of a function in a given interval.
2.	The maximum and minimum values of a function can never occur at the end-points of an interval.	2.	The greatest and the least values of a function may occur at the end-points of the interval.
3.	The maximum value of a function may be <u>less</u> than its minimum value.	3.	The greatest value of a function is always greater than the least value of the function.
4.	A function may not have any maximum or minimum value in a given interval.	4.	A function must have the greatest value and the least value in a given closed interval unless it is a constant-function.

4. Alternative easier approach, if any, in discussing some subtopics.

The usual method for examining a point for maximum or minimum for a function is to apply the method of determining the sign of the 2nd order derivative of the function. But there are several other methods to examine a point for maximum or minimum. These are listed below :

(i) Direct application of the definition of maximum and minimum given in § 3.

Example: Examine the point $x = 0$ for maximum or minimum for the function $f(x) = |x|$.

We note that $f(0) = 0$, $f(0+h) = |h| > 0$

$f(0-h) = |h| > 0$.

$\therefore$ $f(0) <$ both $f(0+h)$ and $f(0-h)$.

Hence, by definition, $x = 0$ is the point of minimum for the function $f(x)$.

(ii) Use of the concept of increasing/decreasing nature of a function.

We note that on the left side of the point of maximum, a function $f(x)$ is increasing and it is decreasing on the right side of this point. Again, on the left side of the point of minimum a function is decreasing and on the right side, it is increasing. This phenomenon helps us to state another rule for examining a given point for maximum or minimum. This rule is stated below :

If the derivative of a function changes sign from positive on the left to negative on the right of a point, then this point is a point of maximum and if the derivative changes sign from negative on the left to positive on the right of the point, then this point is point of minimum.

As an example, if we consider the same function $f(x) = |x|$ for maximum or minimum at $x = 0$, we note that $f'(h) = -1$ if $h < 0$ and $f'(h) = 1$ if $h > 0$. Thus the derivative changes sign from negative on the left to positive on the right of the point $x = 0$. Hence, $x = 0$ is the point of minimum for $f(x) = |x|$.

5. Basic Concept to be emphasised in teaching the topic:

The various concepts of maxima and minima to be emphasised to the teachers in an orientation programme are listed below :

(i) The condition $f'(a) = 0$ for a function $f(x)$ to be a maximum or a minimum at a point $x = a$ is only necessary, but not sufficient. In other words, for a function $f(x)$, $f'(a)$ may be zero, but still $x = a$ may not be a point of maximum or minimum. For example, if $f(x) = x^3$, then $f'(0) = 0$, but it can be shown that $x = 0$ is neither a point of maximum nor a point of minimum.

(ii) The maximum and minimum points of a function are the points where the nature (increasing or decreasing) of a function changes. If the nature does not change at a point, then this point can not be a point of maximum or minimum.

(iii) The maximum and minimum values of a continuous function must occur alternately.

(iv) The maximum value of a function may be less than its minimum value.

(v) A function may have a maximum or a minimum value at a point even if the derivative of the function at this point does not exist.

(vi) The local maximum and local minimum values of a function in a given closed interval can not occur at the end-points of the interval.

6. Analysis of conceptual errors that may be committed by teachers in teaching the topic (in this context, mention gaps and misconception if any, in the text books) which may misguide teachers and students.

In finding maximum and minimum values of a function in a given closed interval, some teachers apply the second order derivative test i.e. they find local maximum and local minimum values and obtain wrong answers. Teachers should remember that maximum or minimum values in a given closed interval are actually the greatest or the least values in the interval.

Some teachers feel that the maximum value of a function is always greater than the minimum value and so they commit conceptual errors in solving a problem of maximum or minimum for a function $f(x)$ which is such that $f^{I}(x) = 0$ for, say, $x = a$ and b $f^{II}(a) > 0$, $f^{II}(b) < 0$, $f(a) = 1$, $f(b) = -16$. Such teachers may conclude that

1 is the minimum and 16 (and not -16) is the maximum value of $f(x)$. As already pointed out that a maximum value of a function may be less than a minimum value and so the maximum value of the above function is not 16 but -16 which is the actual value of $f(b)$ and the minimum value is 1 although $1 > -16$.

If the derivative of a function at a point does not exist, some teachers conclude that the function can not have any extreme value at such a point. But this concept is wrong. For example, $f(x) = |x|$ has a minimum value at $x = 0$ although $f'(0)$ does not exist. The method of determining maximum or minimum values of such functions at such points where the derivatives do not exist will be discussed later.

The definitions of maximum and minimum values of a function given in the textbook of NCERT (pages 158, 161 and 167) are erroneous and may confuse many teachers. These definitions are given below:

1. "A function $f(x)$ is said to have a maximum value in an interval I at x_0 is in I and if $f(x_0) \geq f(x)$ for all x in I. The number $f(x_0)$ is called the maximum value of $f(x)$ in I and x_0 is called a (point of) maximum of $f(x)$ in I.

We can have a similar definition for the minimum value of a function."

2. "Let f be a real function and let x_0 be an interior point in the domain of f. We say that x_0 is a local maximum of f (or a point of local maximum of f or simply, a maximum of f),

If there is an open interval containing xo such that $f(xo) > f(x)$ for every x in that open interval". Similar definition has been given for local minimum of f.

That the above definitions are confusing and defective can be explained as follows :

If we apply definition 1 in finding the maximum (local) value of $f(x) = x^2$ in, say, $0 \leq x \leq 1$, then clearly $f(1) = 1$ is the maximum value, but in fact local maximum can not occur at the end point of an interval.

In case we take I to be an open interval, then definitions 1 & 2 seem to be contradictory (in the first case $f(xo) \geq f(x)$ and in the second case $f(xo) > f(x)$.

The second definition may give a wrong concept to a teacher that a function increasing in an open interval may have a maximum. Besides, this definition does not enable a person to determine maximum values of a function defined in an open interval. Both the definitions do not emphasize the neighbourhood concept of local maximum and local minimum. Within a bigger interval there may be many maxima and many minima, but the above definitions may lead one to conclude that only one maximum or only one minimum may lie in an interval, however large it may be.

The correct definitions of maximum and minimum values of a function are given in § 3.

7. <u>Discussion of some interesting questions that may be asked by the teachers to the Resource Persons</u>

Some of the interesting questions that can be asked by the teachers to the Resource Persons are given below :

(a) Why do we not apply the same method to determine local maximum or local minimum value of a function as well as the greatest or least value of a function ?

(b) In solving physical problems of max. & min., why do we apply the theory of local max. and local min. and not that of absolute max. and absolute min. ?

(c) Can a function have a max. or min. at a point if the derivative at that point does not exist ?

(d) What can you say about the existence of max. or min. value of a function at a point if the first and second order derivatives of the function are zeroes at the point ?

(e) Can a function have two consecutive maxima or two consecutive minima ? If so, what kind of function can it be ?

(f) If the 1st order derivative of a function is a constant and the second and higher order derivatives are all zeroes at a point, what can you say about the existence of maximum, or minimum of the function at that point ?

(g) How can you use the theory of maximum or minimum to determine the sub-intervals of a given interval, in which the function may be increasing and decreasing ?

The Resource Person can explain the above questions in the following way :

(a) The method used to find the greatest and the least values of a function cannot be used to find the max. or min. value of a function because a maximum value of a function may be less than a minimum value, whereas the greatest value of a function in a given closed interval is always greater than its least value.

(b) In absence of a closed interval in a physical problem the theory of finding the greatest and the least values cannot be applied. Besides, for a physical problem there cannot, in general, be more than one maximum and one minimum.

(c) In defining maximum and minimum values of a function at a point, we have considered only the values of the function in a small neighbourhood of the point and not the derivative of the function. Thus a function may have a maximum or minimum value at a point even if the derivative of the function at that point does not exist. The only requirement for a function to have a maximum or a minimum value is that it should be continuous at that point and in a small neighbourhood of the point.

In addition to the necessary condition viz. $f^{I}(a) = 0$ for a function $f(x)$ to have a point of extremum at $x = a$, if $f^{II}(a) = 0$, we can not conclude that $f(x)$ has no max. or min. at $x = a$. This case needs further investigation. It will be explained later that if the second order derivative at a point $x = a$ is zero, we should continue to find higher order derivatives till we find some non-zero derivative of the function at $x = a$. Suppose $f^{n}(a) \neq 0$ for some value of n. Then $f(a)$ is minimum if $f^{n}(a) > 0$ and maximum if $f^{n}(a) < 0$, provided n is even, otherwise there is no maximum or no minimum.

(e) A continuous function can not have two consecutive maxima or two consecutive minima. However, if a function has two consecutive maxima or two consecutive minima, it must have a point of discontinuity in between two maxima and two minima.

(f) If the first order derivative of a function exists but it is non-zero, then the function can not have any maximum or minimum, because the necessary condition for extremum at a point is that the first order derivative of the function must be zero at that point, if the derivative exists.

(g) Suppose a function is defined in $[a, b]$. We find the points of extremum in this interval. Let $x = c$ be the point of maximum and $x = d$ is the point of minimum where $a < c < d < b$. Then the function is increasing in $a < x < c$ and in $d < x < b$ and decreasing in $c < x < d$.

This method can be easily applied in determining the sub-intervals in a given interval, in which a function of the type $f(x) = a \cos(mx + b) \times \sin nx$ is increasing and decreasing, because it is easier to find the

general solution of a trigonometric equation than to find the general solution of a trigonometric inequation.

(Please note that many other similar questions can be asked by the teachers. All such possible questions should be discussed in detail under this section.)

8. Discussion of some enrichment materials on the topic, which the teachers are supposed to know:

The maxima and minima of those functions whose derivatives exist at some points and whose second order derivatives at those points do not vanish have been discussed in the textbook.

We shall now discuss how to find the maximum or minimum value of a function at a point where the derivative does not exist or the second order derivative is zero.

If the derivative of a function $f(x)$ at a point, say $x = a$, does no exist, we shall find $f'(a - h)$ and $f'(a + h)$ where h is a sufficiently small positive number. If $f'(a - h) > 0$ and $f'(a + h) < 0$, then $f(a)$ will be maximum and if $f'(a - h) < 0$ while $f'(a + h) > 0$, then $f(a)$ will be minimum.

Alternatively, we can directly apply the definition to find the extremum values of the function at such a point. We find $f(a - h)$, $f(a)$ and $f(a + h)$. If $f(a)$ is greater than both $f(a - h)$ and $f(a + h)$, then $f(a)$ will be maximum. If, on the other hand, $f(a)$ is less than both $f(a - h)$ and $f(a + h)$, then $f(a)$ will be minimum. In this way,

we can examine the functions

(i) $f(x) = |x|$ for maxima or minima at $x = 0$

(ii) $f(x) = 2 + (x - 2)^{2/3}$ for maxima or minima at $x = 2$.

In case the second order derivative of a function $f(x)$ at a point, sy $x = a$, where the first order derivative at the point is zero, - , then we should find the higher order derivatives. We should continue this process until we get some non-zero derivative of the function.

Let $f^n(a) \neq 0$ for some n.

Then $f(a)$ is maximum if $f^n(a) < 0$ and $f(a)$ is minimum if $f^n(a) > 0$, provided n is an even integer. If n is an odd integer, $f(a)$ is neither a maximum nor a minimum. This result can be easily proved by using Taylor's theorem and expansion of a function. This is beyond the scope of discussion.

The reader should note carefully that all extrema can not be discussed by means of this theorem, even when the derivatives of all orders exist. There exist some functions whose derivatives of all orders may be zero at a point, but still the function may have an extremum at that point.

For example, consider the function :

$$f(x) = e^{-\frac{1}{x^2}} \quad \text{when } x \neq 0$$

$$= 0 \quad \text{when } x = 0.$$

It can be shown that $f^n(0) = 0 \; \forall \; n \in N$. Again both $f(-h)$ and $f(h)$ are greater than $f(0)$ showing that the function has a minimum value

at $x = 0$ although its derivatives of all orders are zero at $x = 0$.

Point of inflexion:

There are certain functions whose first order and second order derivatives at a point are zeroes but the third order derivative is non-zero. Such a point is called the point of inflexion for the function. This is not a point of maximum or minimum. At this point of the curve the tangent is parallel to x - axis, but it cuts the curve at that point. The increasing or decreasing nature of a curve does not change at this point.

Example: $f(x) = x^3$ has a point of inflexion at $x = 0$.

Concavity and Convexity:

A function $f(x)$ differentiable in an interval I is said to be convex if its derivative $f'(x)$ is increasing and concave if its derivative $f'(x)$ is decreasing. Thus, the necessary and sufficient condition for a function $f(x)$ to be convex in I is that $f''(x) \geq 0$ and to be concave in I is that $f''(x) \leq 0 \ \forall x \in I$.

9. Construction of one (or more) intelligent questions (if possible) to test a particular concept and its solution

We discuss below two intelligent problems.

Problem 1:

A polynomial function of degree five has maxima at points $x = b$ and $x = d$ and minima at points $x = a$ and $x = c$. Then which of the following statements are never true ?

A) $c > d > a > b$

B) $b > d > a > c$

C) $c > a > d > b$

D) $b > a > d > c$

E) $a > c > b > d$

F) $d > c > b > a$

G) $a > b > c > d$

H) $d > b > a > c$

I) $c > d > b > a$

J) $d > b > c > a$

K) $c > a > b > d$

L) $d > c > a > b$

M) $b > a > c > d$

N) $a > c > d > b$

O) $b > d > c > a$

P) $a > b > d > c$

We observe that the choices (B), (C), (E), (H), (J), (K), (N) and (O) are never true, for in those cases consecutive maxima and consecutive minima occur, which can not be true for a polynomial function which is continuous for all values of the variables involved.

Problem 2:

Find the derivative of the function

$$y = \sin^{-1} (2x \sqrt{1-x^2}) \quad , \quad 0 < x < 1$$

at $x = \cdot 7$ and $x = \cdot 8$ and hence discuss whether it can have any extremum point between $\cdot 7$ and $\cdot 8$.

We put $x = \sin\theta$

Then $y = \sin^{-1}(2\sin\theta\cos\theta) = \sin^{-1}\sin 2\theta = 2\theta = 2\sin^{-1}x$,

$$\therefore \frac{dy}{dx} = \frac{2}{\sqrt{1-x^2}} \quad \ldots \quad (1)$$

Now, since y is defined in $-\frac{\pi}{2} \leq y \leq \frac{\pi}{2}$,

$$\therefore -\frac{\pi}{2} \leq 2\sin^{-1}x \leq \frac{\pi}{2} \Rightarrow -\frac{\pi}{4} \leq \sin^{-1}x \leq \frac{\pi}{4}$$

$$\Rightarrow -\frac{1}{\sqrt{2}} \leq x \leq \frac{1}{\sqrt{2}} \simeq .71$$

Hence, $\frac{dy}{dx} = \frac{2}{\sqrt{1-x^2}}$ is valid in $-\frac{1}{\sqrt{2}} < x < \frac{1}{\sqrt{2}}$.

Hence, $\left.\frac{dy}{dx}\right|_{x=.7} = \frac{2}{\sqrt{1-.49}} > 0$.

Again, to find $\frac{dy}{dx}$ at $x = .8$, we put

$$x = \cos\theta$$

$$\therefore y = \sin^{-1}(2\cos\theta\sin\theta) = \sin^{-1}\sin 2\theta = 2\theta = 2\cos^{-1}x$$

$\therefore \frac{dy}{dx} = -\frac{2}{\sqrt{1-x^2}}$. Now, $0 \leq \cos^{-1}x \leq \pi$

$$\Rightarrow 0 \leq 2\cos^{-1}x \leq 2\pi \quad \ldots\ldots (2)$$

Also $-\frac{\pi}{2} \leq y = \sin^{-1}\left(2x\sqrt{1-x^2}\right) \leq \frac{\pi}{2} \quad \ldots\ldots (3)$

From (2) and (3) we see that

$y = \cos^{-1}x$ will be in the common interval of (2) and (3)

i.e.

$$0 \leq 2\cos^{-1} x \leq \frac{\pi}{2}$$

$$\Rightarrow 0 \leq \cos^{-1} x \leq \frac{\pi}{4}$$

$$\Rightarrow 1 \geq x \geq \frac{1}{\sqrt{2}}$$

$$\Rightarrow \frac{1}{\sqrt{2}} \leq x \leq 1$$

$\therefore \quad \frac{dy}{dx} = -\frac{2}{\sqrt{(1-x^2)}}$ is valid in $\frac{1}{\sqrt{2}} < x < 1.$

$$\therefore \quad \left.\frac{dy}{dx}\right|_{x=.8} = -\frac{2}{\sqrt{1-.64}} < 0.$$

Thus $\quad \left.\frac{dy}{dx}\right|_{x=.7} > 0$ and $\left.\frac{dy}{dx}\right|_{x=.8} < 0.$

Hence, the given function has a maximum at a point between $x = .7$ and $x = .8.$

10. <u>References (of easy, cheap and easily available bokks).</u>

The following books can be referred to teachers.

(a) Differential and Integral calculus - Piscunov, Peace Publishers, Moscow.

(b) Calculus of one variable - Maron, Mir Publishers, Moscow.

(c) Problems of Mathematical analysis - Demidovich, Peace Publishers, Moscow.

— X —

LIMIT, CONTINUITY AND DERIVATIVES

Prepared by

1. Dr. R. K. Bhattacharya
Department of Mathematics
B. K. C. College
111/2, B. T. Road
Calcutta - 700035

2. Mrs S. Mahapatra
PGT Mathematics
Blue Bells School
New Delhi - 110048

LIMIT, CONTINUITY AND DERIVATIVE.

1. MOTIVATION OF THE TOPIC

Limiting Process is known as the fifth fundamental process in Mathematics. The whole structure of calculus is based on the concept of limit. While the other processes are taught in the earliest stages of schooling, limiting process is taught only at a mature stage. This clearly offers the motivation for learning limit, continuity and derivatives. The motivation for studying limit is intrinsic to the basic development of mathematics itself. Nevertheless the physical world offers motivation for studying the calculus of limit. In nature we find infinitesimal changes taking place in a quantity causing an infinitesimal change in another quantity. Sometimes the ratio of the latter to the former however remains finite. Thus determination of this rate of change of variable quantities assumes greatest importance in the scientific study of social and physical situation. In determining instantaneous velocity we have to take the limit of the ratio $\frac{\Delta s}{\Delta t}$ of two infinitesimal quantities and not the simple ratio $\frac{\Delta x}{\Delta t} = \frac{0}{0}$. The shape of a curve gradually changes with the direction of the tangent. Therefore, the shape of the curve can be defined as the rate of change of the direction of the tangent. Numerous similar examples can be cited to describe the motivation for studying the topic.

The intrinsic mathematical motivation dictates that greatest care has to be adopted in imparting the concept of limit to the first learners. This should however be preceded by an elucidation of what has been termed as Functions. Question may be asked : Is it essential for a function to be expressible by a mathematical form always? This and various other types of functions including functions of several variables are required to be studied through real-life examples. A set approach to defining functions will be meaningful and effective only if pursued through numerous illustrations. Once this has been done, the concept of limit, even $\epsilon - \delta$ concept, can be introduced by considering a simple function and then

painstakingly showing what exactly is meant by the process expressed by the words " tends to ". Let $f(x) = x^2$. Then the concept of $\lim_{x \to 2} f(x)$ is explainable through the following calculation (the intrinsic mathematical motivation in studying the fifth fundamental process is also immediately realisable from this explicit calculation):

x	3	2.5	2.1	2.05	2.01	2.005	2.002	2.001
f(x)	9	6.25	4.41	4.2025	4.0401	4.020025	4.008004	4.004

x	1	1.5	1.8	1.9	1.95	1.99	1.995	1.999	1.9999
f(x)	1	2.25	3.24	3.61	3.8025	3.9601	3.980025	3.996001	3.99960001

From this calculation, the meanings of ϵ, δ, a, l, are explicitly discernible. At this level there is no reason why ϵ - δ definitions cannot be discussed in the classroom pointing out that Right and Left limits are intrinsically implied in the criteria $a - \delta \leq x \leq a + \delta$. The concept of ϵ-neighbourhood steps in easily at this stage consequent to mathematical motivation.

The analytical or limiting criteria of continuity of a function cannot but be initiated by geometrical technique. A graph of a function which has no jump or any other break is a continuous curve. Drawing graphs of continuous and discontinuous curves constitutes the first step in explaining continuity and discontinuity of a function at some points. Thereafter comes the question of understanding continuity in terms of limits. On acquiring a clear conception of limit of a function can one study the other limit known as the Differential Coefficient.

2. <u>Brief Outline of the content</u>

Alongside the brief outline of the content is being mentioned some topics not included in the book. This will give an integrated view of the topic being studied. The object is to define functions by Set Approach and then to acquaint readers with various types of functions. Graphs of some important functions such as Sin x, Cos x, tan x, e^x, etc. are discussed in detail. (Bounded functions, monotone functions are not discussed, whereas inverse functions has been discussed only briefly). Then limit

of a function (not of a sequence) is introduced with examples. Some fundamental facts about limits such as $\underset{x\to a}{Lt}\{f(x)+g(x)\} = \underset{x\to a}{Lt} f(x) + \underset{x\to a}{Lt} g(x)$ etc. has been introduced without proofs to enable readers to use these results in evaluating limits of function. (There always remains a good chance of incorrect use of the results in problem-solving exercises). Some standard limits however have been established by proofs. Examples: $\underset{x\to a}{Lt}\frac{x^n - a^n}{x-a} = na^{n-1}$, n = a positive integer $\underset{\theta\to 0}{Lt}\frac{\sin\theta}{\theta} = 1$. etc. Right and Left limits have been introduced just before discussion on continuity of functions. Removable discontinuity has been discussed but no infinite discontinuity has been mentioned. Examples are inadequate.

Then comes differentiation by first principle, i.e. by evaluating limits directly and then by using formulas. n^{th} order differentiation has been shown. Rules of differentiation of product and quotients of two functions have been shown.

3. <u>Explanation of Technical/Mathematical Terms not properly explained in the textbook</u>

In Calculus, it is not desirable to use results without proofs. Expansions of e^x, sin x, cos x etc. have been freely assumed to construct problems meant for problem-solving exercises (In page 99, the expansion of $\left(1+\frac{n}{x}\right)^{-3/2}$, in Page 80, the expansion of e^x etc). The students will get used to these results without ever seriously going through their mathematical validity. The removable discontinuity, though discussed briefly, need elaboration. The infinite discontinuity of the type $f(x) = e^{-\frac{1}{x-a}}$ at $x = a$ or oscillatory discontinuity $f(x) = \sin\frac{1}{x}$ at $x = 0$, have not been discussed at all. ϵ-neighbourhood has not been discussed even roughly. That not every continuous function is differentiable has been demonstrated by most authors by considering the function $f(x) = |x|$ only (p. 98). Examples like (i) $f(x) = x\sin\frac{1}{x}$, $x \neq 0$, $f(0) = 0$; (ii) If $F(x) = 3 + 2x$ for $-\frac{3}{2} \leq x \leq 0$
$= 3-2x$ for $0 < x \leq 3$

Show that $f(x)$ is continuous at $x = 0$ but $f^1(0)$ does not exist, and some others must be cited to drive the point home. Functions having one-sided derivative or unequal one-sided derivatives (Ex : $f(x) = |x|$) and functions having infinite derivative may be cited to illustrate the complex n... differentiation. Suggested problems for the teacher...

personnel are:

i) Let $f(x) = 0$ when $0 \leq x < \frac{1}{2}$, $f(\frac{1}{2}) = 1$, $f(x) = 2$ when $\frac{1}{2} < x \leq 1$. Show that $f(x)$ is discontinuous at $x = \frac{1}{2}$, $f'(\frac{1}{2})$ exists and its value is infinite;

ii) If $f(x) = x$ for $0 \quad x \quad \frac{1}{2}$, $f(x) = 1 - x$ for $\frac{1}{2} < x \leq 1$, does $f'(\frac{1}{2})$ exist ?

[Solution:

i) $\underset{x \to \frac{1}{2}+0}{\mathrm{Lt}} f(x) = 2$, $\underset{x \to \frac{1}{2}-0}{\mathrm{Lt}} f(x) = 0$

$\therefore$ $f(x)$ is discontinuous at $x = \frac{1}{2}$.

$$R f'(\tfrac{1}{2}) = \underset{h \to 0+0}{\mathrm{Lt}} \frac{f(\frac{1}{2}+h) - f(\frac{1}{2})}{h} = \mathrm{Lt} \frac{2-1}{h} = \infty$$

$$L f'(\tfrac{1}{2}) = \underset{h \to 0-0}{\mathrm{Lt}} \frac{f(\frac{1}{2}+h) - f(\frac{1}{2})}{h} = \mathrm{Lt} \frac{0-1}{-h} = \infty$$

$\therefore$ $f'(\frac{1}{2})$ is infinite.

ii) $$R f'(\tfrac{1}{2}) = \underset{h \to 0+0}{\mathrm{Lt}} \frac{f(\frac{1}{2}+h) - f(\frac{1}{2})}{h} = \underset{h \to 0+0}{\mathrm{Lt}} \frac{1 - \frac{1}{2} - h - \frac{1}{2}}{h} = -1$$

$$L f'(\tfrac{1}{2}) = \underset{h \to 0-0}{\mathrm{Lt}} \frac{f(\frac{1}{2}+h) - f(\frac{1}{2})}{h} = \mathrm{Lt} \frac{\frac{1}{2} - h - \frac{1}{2}}{-h} = 1$$

$\therefore$ $f'(\frac{1}{2})$ does not exist.]

Although the exponential 'e' has been freely used in the book (Pages 80 - 83) $\underset{n \to \infty}{\mathrm{Lt}} (1 + \frac{1}{n})^n$ has not been mentioned. The expansion of $\log(1+x)$ has been assumed and freely used to solve problems. The question of McLaurin's conditions of validity need be emphasised even when the assumptions are being made (P. 82).

4. <u>Some Conceptual errors</u>

a) The limit $\underset{x \to 0}{\mathrm{Lt}} \frac{e^x - 1}{x} = 1$ in page 81, has been evaluated by taking recourse to the McLaurin's expansion of e^x, completely forgetting that the same limiting value was employed in McLaurin's expansion where $\frac{d}{dx}(e^x)$ was involved. $\frac{d}{dx}(e^x) = \underset{h \to 0}{\mathrm{Lt}} \frac{e^{x+h} - e^x}{h} = e^x \underset{h \to 0}{\mathrm{Lt}} \frac{e^h - 1}{h}$

$= e^x . 1 \quad e^x$.

Thus there is clearly a cycling type of error which is [illegible] ...ly [illegible] error.

b) Similar error may occur if $\underset{x \to 0}{Lt} \frac{\sin x}{x}$ be evaluated by by L' Hospital's Rule; for $\frac{d}{dx}(\sin x) = \cos x$ has already been derived by using the Limit $\underset{x \to 0}{Lt} \frac{\sin x}{x} = 1$

[$\because \underset{x \to 0}{Lt} \frac{\sin x}{x} = \underset{x \to 0}{Lt} \frac{\cos x}{1} = 1$ is wrong.]

5. Some Enrichment Materials for Resource Personnel

1) Fundamental Theorems on Limits cannot be applied mechanically. For example $\underset{x \to a}{Lt} \left\{ f(x) + g(x) \right\} = \underset{x \to a}{Lt} f(x) + \underset{x \to a}{Lt} g(x)$ is not applicable for evaluating (taking term by term limits)

$$\underset{x \to 0}{Lt} \left(\frac{x^2}{1+x^2} + \frac{x^2}{(1+x^2)^2} + \dots\dots\dots \right).$$

An erroneous application of the theorem gives the limit

$$= \underset{x \to 0}{Lt} \frac{x^2}{1+x^2} + \underset{x \to 0}{Lt} \frac{x^2}{(1+x^2)^2} + \dots$$

$$= 0 + 0 + \dots\dots\dots = 0$$

Correct Limit $= \underset{x \to 0}{Lt} x^2 \left[\frac{\frac{1}{1+x^2}}{1 - \frac{1}{1+x^2}} \right]$

$$= \underset{x \to 0}{Lt} x^2 \cdot \frac{1}{x^2} = 1$$

2) The theorem $\underset{x \to a}{Lt} f(x)\, g(x) = \underset{x \to a}{Lt} f(x)\, \underset{x \to a}{Lt} g(x)$ is valid only when $\underset{x \to a}{Lt} f(x)$ and $\underset{x \to a}{Lt} g(x)$ both exist. It would be wrong to write

$$\underset{x \to 0}{Lt}\, x \sin \frac{1}{x} = \underset{x \to 0}{Lt}\, x \cdot \underset{x \to 0}{Lt} \sin \frac{1}{x} = 0 \times \underset{x \to 0}{Lt} \sin \frac{1}{x} = 0.$$

3) By a function f from (or on) a set X to (or into) a set Y we mean a rule that assigns to each x in X a unique element of f (x) in Y. The collection G of pairs of the form $\langle x, f(x) \rangle$ in X x Y is called the GRAPH of the FUNCTION f. A subset G of X x Y is the Graph of a function on X if and only if for each $x \in X$ there is a unique pair in G whose first element is x.

Since a function is determined by its graph, it is advisable to define the graph in terms of sets. In the book only function but not graph of the function has been defined in terms of sets.

4) Limit of a sequence may be considered taking a simple case. The word "sequence" is not in the Book.

6. <u>Some Suggestions and Problems based on Alternative Approach</u>

1) L'Hospital's Rule is not mentioned in the Book. The proof requires a knowledge of Cauchy's Mean Value Theorem. Since Lagrange's Meann Value Theorem is already in the Book, these two theorems may be described to get the students acquainted with Application of L'Hospital's Rule in finding limits. Problems suggested:

a) Find a, b such that

$$\underset{x \to 0}{\mathrm{Lt}} \frac{x\,(1+a \cos x) - b \sin x}{x^3} = 1\text{ , (apply L.H. Rule)}$$

b) Evaluate: $\underset{x \to \frac{\pi}{2}}{\mathrm{Lt}} (\tan x)^{\cos x}$, $\underset{x \to 0}{\mathrm{Lt}}\, x^{\sin x}$, $\underset{x \to 0}{\mathrm{Lt}} (1-\sin$

$$\underset{x \to 0}{\mathrm{Lt}} \left(\operatorname{cosec} x - \frac{1}{x} \right)\text{, etc. (apply L.H. Rule)}$$

2) Removable discontinuity has been defined as (P.91). "If a is a point of discontinuity such that $\underset{n \to a}{\mathrm{Lt}} f(x)$ exists, then by changing the value of f at a, we can make it continuous at a. We say then that a is a removable discontinuity of f ". This definition should be replaced by : " If $f(a+0) = f(a-0) \neq f(a)$, or $f(a)$ is not defined, then $f(x)$ is said to have a removable discontinuity at $x = a$ ".

An example: $f(x) = (x^2 - a^2) / (x - a)$ has a removable discontinuity at $x = a$, for $f(a)$ is undefined here, though $\underset{x \to a}{\mathrm{Lt}} f(x)$ exists and $\underset{x \to a}{\mathrm{Lt}} f(x) = \underset{x \to a}{\mathrm{Lt}} \dfrac{(x+a)(x-a)}{(x-a)} = 2a$,

Another example : Let $f(x) = e^{-(x-a)^{-2}}$, $x \neq a$

$= 1$, $x = a$

Here $\lim_{x \to a} f(x) = 0$ and $f(a) = 1$

So $x = a$ is a removable discontinuity of $f(x)$.

Example: $f(x) = [x] + [-x]$ at $x = 0$, has a simple non-removable discontinuity at $x = 0$.

3) In Page 90, a sentence runs as : " Let a = 1. Then the nearby points can be either > 1 or < 1, " meaning probably the points in the neighbourhood. The 'neighbourhood' concept should be used by writing the word NEIGHBOURHOOD in a straightforward manner.

4) The Word ' infinitesimal' has to be used and its meaning explained.

5) In Page 83, a sentence runs as " We shall later prove the stronger result that every polynomial function is continuous at every point". The word "Stronger" may be replaced by the word "general".

6) In the case of derivatives of inverse Trigonometric functions care must be taken to explain the existence or otherwise of derivatives at the end points:

Example:- $\frac{d}{dx}(\sec^{-1} x) = \frac{1}{|x|\sqrt{x^2 - 1}}$ for $|x| > 1$

and for $0 < \sec^{-1} x < \pi$.

The derivative of $\sec^{-1} x$ does not exist at the end points although $\sec^{-1} x$ is defined at the end points.

7. <u>Basic Concepts to be Emphasized in Teaching the topic</u>

1) Since the students have some good knowledge of set theory and since FUNCTION of a real variable has been defined in terms of Set, it would be advisable to impress upon the students the set theoretic definition of GRAPH and other terms. This will help them to pursue mathematical analysis at the Degree Level.

2) A thorough discussion on the continuity of a function at the end points of the domain should be made. A function defined in the closed interval $[a, b]$ is not continuous at the end points for continuity at 'a' the left limit will exist only if the domain be extended beyond 'a' to the left. Similarly for continuity at b the right limit will exist if the domain be extended beyond 'b' to the right. This question will also help understand conditions of Rolle's Theorem properly.

3) It is to be impressed that computation of derivative is actually finding limits. This fact should not be lost sight of while computing derivative mechanically by using formulas.

8. Alternative Method of Solving some Problems

1). In Page 68, $\underset{x \to 0}{\mathrm{Lt}}$ $\sqrt{\text{[redacted]}} - 1$ [redacted] can be solved easily as follows:

without taking recourse to using the limit $\underset{x \to a}{\mathrm{Lt}} \dfrac{x^n - a^n}{x - a}$.

$$\text{The given limit} = \underset{x \to 0}{\mathrm{Lt}} \frac{1 + x - 1}{x(\sqrt{1+x} + 1)} = \frac{1}{2}.$$

2) In Page 124, $y = b \tan^{-1}\left\{\frac{x}{a} + \tan^{-1}\frac{y}{x}\right\}$

should be differentiated from the step

$\tan y = b\left(\frac{x}{a} + \tan^{-1}\frac{y}{x}\right)$ and not as done in the book.

References

1. Differential Calculus - B.C. Das & B.N.Mukherjee (Calcutta)
2. Differential Calculus - K.C. Maity & R.K.Ghosh (Calcutta)
3. Real Analysis - H.L. Royden, Maxwell McMillan International Edition, New Delhi
4. Principles of Real Analysis - S. C. Malik (New Delhi)

ROLLES THEOREM AND LAGRANGE'S MEAN VALUE THEOREM

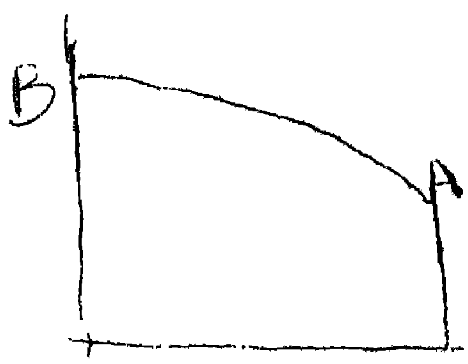
Prepared by

1. Dr (Mrs) Mira Sarkar
Lady Brabourne College
Department of Mathematics
P 1/2, Suhrawardy Avenue
Calcutta - 700001

2. Mrs S. N. Shantha Ramana
C. G. H. S. School
Hotler's Street
Madras - 600112

ROLLE'S THEOREM AND LAGRANGE'S MEAN-VALUE THEOREM

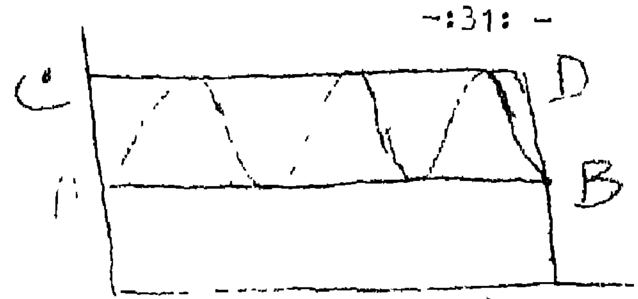
1) Motivation of the topic



If a stone is thrown from a point A to hit a vertical wall at the point B situated at a height different from A, is it ever possible that its velocity is horizontal at some point of the parabolic path it describes? What happens if A and B are at the same height ?

Is it ever possible that the velocity is never horizontal in any of the above situations? Or is it possible that the direction of its velocity at a point C is ever parallel to the line AB, even if A and B are at different heights and C lies between A and B ?

Or, if on the occasion of a festival, a decorator bends a thin metal sheet a number of times (as is shown in the figure) so that it touches a horizontal bar CD on the top of the gate placed between its



two vertical pillars, where (i) no overlapping of the sheet takes place between A and B.

(ii) the portion between A and B is a smooth curve without any break or sharp bend, should there be, any restriction on the heights of A and B ?

Rolle's Theorem and Mean Value Theorem help us to solve these problems.

In the first example, if A and B are at the same height, Rolle's Theorem says that the velocity of the stone will be horizontal at least once before it hits the wall. But if they are at different heights, the direction of velocity will be parallel to the line AB at least once before the stone reaches B, a result given by the Mean Value Theorem. The possibility of it being zero cannot be ensured now.

In the second example, these theorems suggest that A and B can lie at any height.

Many other situations may arise in our daily lives, which can be analysed best by Rolle's and Mean Value Theorems.

2) Brief Outline of the Content.

Statement of Rolle's Theorem. If a function $f(x)$ defined over the closed interval $[a, b]$ is such that

(i) it is continuous in $[a, b]$

(ii) $f'(x)$ exists in the open interval $]a, b[$ and

(iii) $f(a) = f(b)$

then there exists at least one point c,

$a < c < b$, such that f' ...

<u>The geometrical significance of it is as follows:</u>

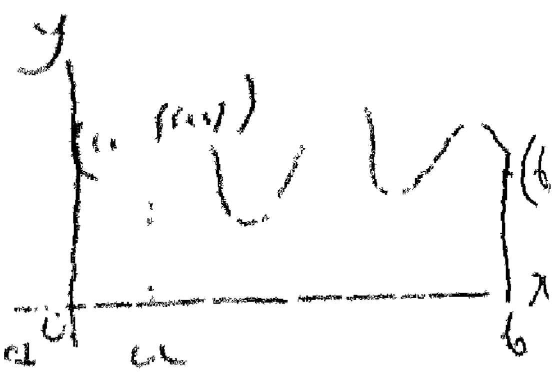
If the portion of the graph of the function $y = f(x)$ between the ordinates $x = a$ and $x = b$ is such that

(i) it is a continuous curve from the point $(a, f(a))$ to $(b, f(b))$

(ii) it has a clear tangent everywhere between those two points,

and (iii) the line joining these points is parallel to the x-axis,

then there must exist <u>at least</u> one point on the curve between these points, where the tangent is parallel to the x-axis.



<u>Statement of the Mean Value Theorem</u>

If the function $f(x)$ defined over $[a, b]$ is such that

(i) it is continuous in $[a, b]$

(ii) derivable in $]a, b[$,

then there will exist at least one point c, $a < c < b$, such that

$$f'(c) = \frac{f(b) - f(a)}{b - a}$$

This is proved by constructing a function $\phi(x) = f(x) + Ax$, where A is some constant. Details of it is given in the next article.

Geometrical significance

The hypothesis of the theorem implies the conditions (i) and (ii) of Rolle's Theorem.

Then it implies that there will exist at least one point between the points $\left(a, f(a)\right)$ and $\left(b, f(b)\right)$, the tangent at which is parallel to the chord joining these points.

3. Explanation of topics not included in the book.

Mean Value theorem has not been proved in the NCERT book, though it is included in the syllabus. This is given as follows :

Proof of Mean Value Theorem :

Let $\phi(x) = f(x) + Ax$ where A is a constant so chosen that $\phi(a) = \phi(b)$. That is $f(a) + A(a) = f(b) + Ab$

$$\Rightarrow A = -\frac{[f(b) - f(a)]}{b - a}$$

But this choice of A makes $\phi(x)$ satisfy all the conditions of Rolle's Theorem.

So $\phi'(c) = 0$ for some c in $]a, b[$

Therefore, $f'(c) + A = 0 \Rightarrow A = -f'(c)$.

Equating the values of A, the proof is completed.

4. Analysis of possible conceptual errors.

1. The conditions of Rolle's Theorem may be misunderstood by some as necessary. The following examples will show that they are only sufficient.

Example (i) If $f(x) = \frac{1}{x} + \frac{1}{1-x}$, it is neither continuous nor differentiable in $[0, 1]$. In fact both $f(0)$ and $f(1)$ are undefined.

But $f'(x) = \frac{2x-1}{x^2(1-x)^2}$, which exists in $]0, 1[$. So the conditions of Rolle's Theorem are not all satisfied here.

But still $f'(x) = 0$ at $x = \frac{1}{2}$, a point in the open interval $]0, 1[$.

Example (ii) If $f(x) = |x|$, it is continuous in $[-1, 1]$; but $f'(0)$ does not exist while $f(-1) = f(1)$. Here also the conditions of Rolle's Theorem are not all satisfied in the interval. This function, of course, does not have a zero derivative anywhere, so that for it, the conclusion of the theorem also is not true.

(c) Again, existence of only one point in the open interval $]a, b[$ may be thought to be necessary in the case of both these theorems. The following examples will confirm that this is not so.

Example (i) If $f(x) = x(x-1)^2 (x-2)^2$, $a = 0$, $b = 2$, it is continuous in $[a, b]$.

$$\begin{aligned} f'(x) &= (x-1)^2 (x-2)^2 + 2x(x-1)(x-2)^2 + 2x(x-1)^2(x-2) \\ &= (x-1)(x-2)\left[(x-1)(x-2) + 2x(x-2) + 2x(x-1)\right] \\ &= (x-1)(x-2)(5x^2 - 9x + 2) \end{aligned}$$

So $f'(x)$ exists in $]0, 2[$ and vanishes at $x = .26, 1, 1.54$ and 2. Of these, the first three points lie in $]0, 2[$.

(ii) If $f(x) = |x|$, $a = 0$ and $b = 1$, it is continuous in $[a, b]$ while $f'(x)$ exists in $]0, 1[$ because $f'(x) = 1, x > 0$
$= -1, x < 0$

$\frac{f(1) - f(0)}{1 - 0} = 1 = f'(x)$ for all x in $]0, 1[$.

So the 'c' of Mean Value Theorem can be anywhere in $]a, b[$.

5. Some Probable questions on the topic.

1. If $f(x)$ is continuous in only the open interval $]a, b[$, will the theorem hold ?

Ans: No. $f'(c) = 0$ may not always be true.

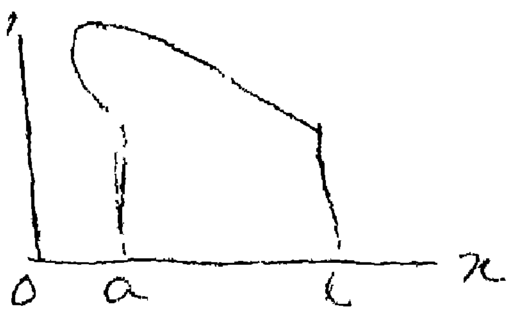
This can be illustrated by the following example:

(i) If $f(x) = \frac{1}{x}$ $a = 0$, $b = 1$, it is continuous and differentiable in $]a, b[$. But no 'c' exists for which $f'(c) = 0$.

(ii) If $f(a) \neq f(b)$, there may not be any c between $]a, b[$ such that $f'(c) = 0$.

For example: if $f(x) = x^2$, $a = 1$, $b = 2$, it is continuous and differentiable in $[1, 2]$ but there is no 'c' in $]1, 2[$ for which $f'(c) = 0$, for here $f(1) \neq f(2)$.

(iii) The curve on the left satisifes all the conditions of Rolle's Theorem but no tangent can be drawn parallel to the x-axis. Why ?



Answer: Because this is not the graph of a function. (it is not single valued.)

(iv) In proving the Mean Value Theorem, why should $\phi(x)$ be taken in that particular form.

ANS: This may be only because it gives the desired result. In fact, for proving Mean Value Theorems of higher order, say the one of order n, $\phi(x)$ is taken as

$$f(x) + (b-x)f'(x) + \frac{(b-x)^2}{2!}f''(x) + \cdots + \frac{(b-x)^{n-1}}{(n-1)!}f^{n-1}(x) + (b-x)^n A,$$

where A is a constant so chosen that

$$\phi(b) = \phi(a).$$

From this the final result, namely,

$$f(b) = f(a) + (b-a)f'(a) + \frac{(b-a)^2}{2!}f''(a) + \cdots + \frac{(b-a)^{n-1}}{(n-1)!}f^{n-1}(a) + \frac{(b-a)^n}{n!}f^n(c), \quad a < c < b.$$

can be derived.

Necessary modifications are to be made depending on the order of the theorem.

6. Enrichment Material.

In the Mean Value Theorem, if $b = a+h$ the theorem reads :

$$\frac{f(a+h) - f(a)}{h} = f'(c), \text{ where } a < c < a+h.$$

So 'c' can be replaced by $a+\theta h$, where $0 < \theta < 1$.

Hence, $f(a+h) = f(a) + h\, f'(a+\theta h)$, $0 < \theta < 1$ which is an equivalent well-known form of Mean Value Theorem.

APPLICATION OF DERIVATIVES

Prepared by

1. Dr (Mrs) Mira Sarkar
Lady Brabourne College
Department of Mathematics
P1/2, Suhrawardy Avenue
Calcutta - 700017

2. Mrs S. N. Shantha Ramana
C. G. H. S. School
Rotler's Street
Madras - 600112

Application of Derivatives

1) **Motivation of the topic**

In our daily life, we come across situations, some of which are enumerated in the following:

i) The height of a person increases upto a certain age and then remains constant;

ii) The physical strength of a man increases upto a certain age and starts decreasing;

iii) If a particle is projected upwards, it reaches a certain height and then falls downwards till it reaches the ground;

iv) When a train starts from a station, its velocity gradually increases, after which it moves with constant speed for sometime and then gradually slows down and stops;

In all these we find that a particular quantity increases, decreases or remains constant as some other quantity changes. In the first two cases, the age of the person is the second quantity which produces changes in the first, while in the last two, time is the second producing effects on the first.

These changes in the behaviour patterns of the first quantity effected by changes in the second can all be explained if the first is looked upon as a function of the second and its derivatives with respect to the second are carefully analysed at various points.

2) **Brief Outline of the Content**

Derivatives have been applied in the following cases:

i) To determine the equations of tangents and normals to a curve at a point, since $\frac{dy}{dx}$ measures the slope of the tangent; the tangent at (x_0, y_0) to the curve $y = f(x)$ has equation

$$y - y_0 = f'(x_0)(x - x_0)$$

while the normal is given by

$$y - y_0 = -\frac{1}{f'(x_0)}(x - x_0).$$

ii) To determine whether a function is increasing or decreasing.

If $f'(c) > 0$, the function $f(x)$ increases at $x = c$, i.e. $f(c-h) < f(c) < f(c+h)$ for small positive values of h. Similarly, if $f'(c) < 0$, $f(x)$ decreases at $x = c$ i.e. $f(c-h) > f(c) > f(c+h)$ for small positive values of h.

iii) To determine the rate of change of a function for any particular value of its variable.

Thus in problems of mechanics, $\frac{dx}{dt}$ measures the velocity, i.e. the rate of displacement at time 't', while $\frac{d}{dt}\left(\frac{dx}{dt}\right) = \frac{d^2x}{dt^2}$, the acceleration, where x measures the distance described by a particle in time 't'. So the velocity increases or decreases at a point according as $\frac{d^2x}{dt^2} \gtrless 0$ there.

iv) In tracing a curve, the sign of $\frac{dy}{dx}$ at a point determines the upward or downward trend of the curve at the point. It also indicates whether the curve has a tangent parallel to any of the coordinate axes.

3. <u>Explanations of topics not mentioned in the book</u>

i) The differential of a function $y = f(x)$ is defined as $\underline{dy} = f'(x) \Delta x$, where Δx is a small change in the independent variable x. If $f(x) = x$, in particular, this gives $dx = 1. \Delta x$. In the NCERT book (Page 184), it is only stated that dx is also used to denote a small increment Δx, but no reason has been given for it.

Derivative of y with respect to x is, therefore, $\frac{dy}{dx} = \frac{\text{differential of } y}{\text{differential of}}$

i.e. $\frac{d}{dx}(y) = \frac{dy}{dx}$

This is not indicated in the book.

Similarly, the relation $dy = f'(x)\, dx$ explains why $f'(x)$ is called the differential coefficient. This also is not included in the book.

ii) Equation of a tangent is given as $y - y_0 = f'(x_0)(x - x_0)$ ... (1) (Page 152, NCERT book). But this determines the tangent only if $f'(x_0)$ is finite, i.e. when the tangent is not parallel to the y - axis. Since curves do have tangents parallel to the y - axis, the equation (1) should be rewritten as

$$y - y_0 = \frac{1}{\frac{1}{f'(x_0)}}(x - x_0)$$

or, $(y - y_0)\frac{1}{f'(x_0)} = (x - x_0)$, so that if the tangent is parallel to the y axis, when $\frac{1}{f'(x_0)} = 0$, it has equation $x - x_0 = 0$. This is in line with equation of a line parallel to the y - axis in coordinate Geometry.

For the equation of the normal, however, the case $f'(x_0) = 0$ has been considered.

iii) In dealing with curve-tracing (NCERT book page 1[illegible]6 etc.), symmetry about only the origin has been considered, while many of the standard curves are symmetrical about the coordinate axes, the tracing of which is not very difficult. [illegible] whether a curve is closed or has an infinite branch, has also not been discussed.

<u>Examples:</u> i) $x^2 + y^2 = a^2$ (ii) $\frac{x^2}{a^2} + \frac{y^2}{b^2} = 1$

iii) $y^2 = 4ax$ etc.

For this, we add the following :

For the curve (i) $y = \pm\sqrt{a^2 - x^2}$

Therefore $|x| \not> |a|$ i.e. $-|a| < x < |a|$

Again for any admissible value of x, y has two values, equal in magnitude and opposite in sign. This means that the curve is symmetrical about the x - axis.

Similarly, from the same equation we obtain $x = \pm\sqrt{a^2 - y^2}$, so that the curve is symmetrical about the y - axis and $-|a| < y < |a|$

$x = 0$ iff $y = \pm a$ and $y = 0$ iff $x = \pm a$, so this is a closed curve symmetrical about both the coordinate axes and contained

wholly within the lines $x = \pm a$ and $y = \pm a$.

The curve (ii) will similarly be a closed one having symmetry about both the coordinate axes.

For the curve (iii), (o, o) is a point on it. We have $x = \frac{y^2}{4a}$, so $x \gtrless 0$ according as $a \gtrless 0$. Hence for a particular sign of a, this curve lies entirely on one side of the y-axis.

Again, $y = \pm 2\sqrt{ax}$ implies, as above, that the curve has symmetry about the x - axis. y has a real value for all these values of x. Hence, the curve extends to infinity.

4. Analysis of possible conceptual errors

i) The fact that $f'(x) > 0$ implies increasing nature of f (x), may be misunderstood as a necessary condition. But there are functions which increase at a point, even though $f'(x) \ngtr 0$ there.

Example 1: If $f(x) = x^3$, we find $f'(x) = 3x^2$ which vanishes at $x = 0$. But f (x) increases at $x = 0$.

Example 2: If $f(x) = \sqrt[3]{x}$, $f'(x) = \frac{1}{3} x^{-\frac{2}{3}}$, which does not exist at $x = 0$. There also f (x) increases at $x = 0$. Similarly, if $f(x) = -x^3$, f (x) decreases at $x = 0$, though $f'(0) = 0$.

ii) Though dy is taken to measure an approximate change in y, it never measures the actual change.

5. Discussion of questions that may be raised from the topic

i) If $\frac{dy}{dx}$ is neither positive nor negative, what happens?

Ans: The point where this occurs has a stationary ordinate, as it can neither increase nor decrease, if $\frac{dy}{dx}$ exists there. In fact $\frac{dy}{dx}$ must then be zero. These are the points where the function may have an extremum or a point of inflexion (depending on the values of the higher order derivatives). Geometrically the tangent to the curve at this point must be parallel to the x - axis.

If however, $\frac{dy}{dx}$ does not exist at the point no tangent can be drawn to the curve $y = f(x)$ at the point. But the nature of the curve before or after the point cannot be ascertained.

ii) If $\frac{dy}{dx}$ is non-existent at a point but undergoes a change of sign as x passes through this point, what happens ?

Ans: The function will have an extremum at this point. It is maximum or minimum according as the change of sign is from positive to negative or from negative to positive.

The function of course must have a definite value at this point now.

THEORY OF COMPLEX NUMBERS

Prepared by

i) Prof. A. Chakraborty
Department of Mathematics
Jadavpur University
Calcutta - 700 032

ii) Dr. G. P. Mukherjee
SCERT
25/3, Ballygunge Circular Road
Calcutta - 700 019

iii) Dr. (Mrs.) S. Mukherjee
Lady Brabourne College
Department of Mathematics
P1/2, Suhrawardy Avenue
[illegible] 700 017

iv) Sri N. L. Lahiri
B. K. C. College
Department of Mathematics
111/2, B.T. Road
Calcutta - 700 035

THEORY OF COMPLEX NUMBERS

1. Motivation of the Topic

Students have learned in their previous classes how to find the solutions of linear equations such as $3x + 9 = 0$, the simultaneous equations $\left.\begin{aligned} 2x + 5y &= 7 \\ 3x + 5y &= 8 \end{aligned}\right\}$ or even the quadratic equations of the type $x^2 - 5x + 6 = 0$ (giving solutions $x = 2, 3$). However, in the case of equations of the type $x^2 + 25 = 0$, $x^2 + 16 = 0$ or $x^2 + 1 = 0$, $x^3 + 1 = 0$ they are unable to find the complete solution (or solutions) at all as real numbers. By solving $x^2 + 1 = 0$, we have

$$x^2 = -1$$

$$\Rightarrow x = \pm\sqrt{-1}$$

Students are not familiar with the square root of a negative. In fact it was Leonard Euler (1707–83) who, while attempting to find the solution of $x^2+1=0$, introduced the symbol $i = \sqrt{-1}$. With the help of this symbol it became possible to express the solutions of the equation as $x = i, -i$. The introduction of this symbol 'i' has thus led to the development of the theory of complex numbers.

2. Brief Outline of the contents

For any two real numbers a and b we can form a number $a + ib$ known as a complex number. The set of all complex numbers is denoted by c. Usually a complex number is denoted by $Z = a + ib$, a is called the real part, denoted by Re (Z) and b the imaginary part of Z denoted by Im (Z). Carl Fredrich Gauss (1777–1855) introduced the term "complex numbers".

If in Z, [illegible] then $Z \in R$, the set of real numbers. When $a = 0$, then $Z = ib$ which is a purely imaginary number. If both $a = 0$ and $b = 0$, then the number $Z = 0 + i0 = 0$.

It is evident that any number, whether purely imaginary or purely real, can be considered as a complex number : Thus $R \subset C$. The conjugate of $Z = a + ib$ is defined as $\bar{Z} = a - ib$.

A complex number $Z = x + iy$ may also be represented geometrically in what is known as the Argand diagram as a point (x, y) referred to the x-axis as the real axis and the y axis as the imaginary axis. Thus the complex number may be considered as an ordered pair. A polar representation may also be obtained by putting $x = r \cos \theta$, $y = r \sin \theta$ so that $Z = r(\cos \theta + i \sin \theta)$

$$= re^{i\theta}.$$

where r is known as the modulus of Z and is denoted by $|Z|$, and θ is known as the argument of the number Z and is denoted by arg. Z. It is evident that $r = \sqrt{x^2 + y^2}$ [illegible] $\theta = \tan^{-1} (y/x)$ i.e.

$$|Z| = \sqrt{x^2 + y^2} \quad \text{and arg. } Z = \tan^{-1} (y/x).$$

For $|Z|$ we take the positive value of $\sqrt{x^2 + y^2}$. The value of θ satisfying $-\pi < \theta \leq \pi$ is known as the principal argument and is denoted by $\arg_p Z$. Every complex number represents a unique point on the plane and every point on the plane corresponds to a unique complex number. The plane is called the <u>complex plane</u> or the <u>Argand plane</u> after the name of the French mathematician J. R. Argand (1768 - 1822).

3. Explanation of technical / mathematical terms not explained in the text book.

While introducing the polar representation of complex numbers for the polar coordinates (r, θ) the angle θ was such that $0 \leq \theta < 2\pi$ (art 2.3 of NCERT text book). However, further on in the text the value of θ for the principal value of arg Z is such that $-\pi < \theta \leq \pi$. The ranges for the values of θ in either case do not differ. The choice of arg Z, the principal value such that $-\pi < \theta \leq \pi$ may be treated as conventional.

4. Alternative easier approach, if any, in discussing some sub-topics.

1. Instead of using x, y as independent variables we may introduce Z and $\bar{Z}$ as two independent variables. All results may be obtained in terms of Z and $\bar{Z}$.

2. To find the square roots of a complex number of the form a + ib instead of applying De Moivre's theorem we can extract the square root by mental arthmetic only. The method will be clear from the following examples.

Example: Ex. 1 Find the square root of

$$7 + 24i$$

Ans:

$$7 + 24i = 7 + 2.4.3i = 4^2 - 3^2 + 2.4.3i$$

$$= 4^2 + (3i)^2 + 2.4.3i$$

$$= (4 + 3i)^2$$

$$\therefore \sqrt{7 + 24i} = \pm (4 + 3i)$$

Ex. 2. Find the square root of

$$a + 2i\sqrt{a+1}\ , \text{ where } a \text{ is real.}$$

Ans: $a + 2i\sqrt{a+1}$

$$= a + 1 - 1 + 2i\sqrt{a+1}$$

$$= (\sqrt{a+1})^2 + i^2 + 2i\sqrt{a+1}$$

$$= (\sqrt{a+1} + i)^2$$

$$\therefore \sqrt{a + 2i\sqrt{a+1}} = \pm\ (\sqrt{a+1} + i)$$

Another method of extracting the square root:

Find the square root of $7 + 24i$

Let $\sqrt{7 + 24i} = x + iy$

Squaring,

$$7 + 24i = (x^2 - y^2) + 2i\ xy.$$

Equating the real and imaginary parts

$$x^2 - y^2 = 7 \ldots (1) \quad \text{and} \quad xy = 12.$$

$$\therefore (x^2 + y^2)^2 = (x^2 - y^2)^2 + 4x^2y^2$$

$$= 7^2 + (24)^2 = 49 + 576 = 625$$

$$\therefore x^2 + y^2 = +25 \ldots (2) \qquad (\text{Since both } x^2 \text{ and } y^2 \text{ are } +ve)$$

Adding (1) and (2) $2x^2 = 32$ $\quad\therefore x^2 = 16$

$\Rightarrow x = \pm 4$

Subtracting, $2y^2 = 18$ $\quad\therefore y^2 = 9$

$\therefore y = \pm 3$

But $xy = 12$.
So take x, y to have the same sign.
Either $x = 4$, $y = 3$ or $x = -4$, $y = -3$.

$$\therefore \sqrt{7+24i} = \pm\ (4 + 3i).$$

5. <u>Basic concept to be emphasized in teaching the topic:</u>

The various concepts of complex numbers to be emphasized to the teachers in an orientation programme are listed below :

(i) In the expression $x + iy$ the sign + does not indicate addition as understood by the usual manner. In the expression $x + iy$ means is that x real units are combined in order by y imaginary units to get a complex number $Z = x + iy$.

(ii) If $a + ib = c + id$, where a, b, c, d are real numbers, then $a = c$ and $b = d$.

(iii) If $a + ib = 0$ where a and b are real numbers, then $a = 0$, $b = 0$.

(iv) The modulus of complex number $Z = x + iy$ should be taken to be the positive value of $\sqrt{x^2 + y^2}$.

(v) (a) The modulus of the sum of any number of complex numbers is less than or equal to the sum of their moduli.

(b) The modulus of the difference of two complex numbers is greater than or equal to the difference of their moduli.

(c) The modulus of the product of any number of complex numbers is equal to the product of their moduli and the amplitude of the product is equal to the sum of their amplitudes.

(d) The moduli of the quotient of two complex numbers is equal to the quotient of their moduli (provided the denominator $\neq 0$) and the amplitude of the quotient is equal to the difference of their amplitudes.

(vi) The principal value of the argument θ should be as $-\pi < \theta \leq \pi$. Unless otherwise stated we mean the principal value of the argument.

If $z = x + iy$ is a complex number, then $r = +\sqrt{x^2 + y^2}$ and the amplitude θ should be determined by the equations $\cos\theta = \frac{x}{r}$ and $\sin\theta = \frac{y}{r}$. If we consider $\theta = \tan^{-1}(y/_x)$ only, sometimes we may have a wrong answer.

(vii) The numbers i and $-i$ have a unique characteristic that each is simultaneously additive inverse, multiplicative inverse and the conjugate of the other.

(viii) Any integral power of i is either $+i$ or $-i$ or $+1$ or -1.

(ix) Any integral power of ω must have the value either $1, \omega$ or ω^2.

(x) The real numbers are ordered. But the complex numbers are not ordered i.e. if z_1 and z_2 be two complex numbers, we cannot say whether $z_1 > z_2$ or, $z_1 < z_2$ (the case of equality has been discussed in (iii)).

(xi) There is a one-one correspondence between ordered pairs in the Argand plane and complex numbers, but with a polar representation the one-one correspondence between the points (r, θ) in the polar plane and the complex numbers does not exist unless the principal values of θ are considered.

6. Analysis of conceptual errors [illegible] committed by [illegible] misconceptions, in teaching the topic [illegible] if any, in the textbook [illegible]

a) In finding the modulus of a complex number by the formula $|Z| = +\sqrt{x^2+y^2}$, sometimes they may [illegible] answer. This will be clear from the following example. Suppose we are to find the modulus of $1 + i \tan 3/5\pi$.

If we take $Z = 1 + i \tan 3/5\pi$, then $x = 1$, $y = \tan 3/5\pi$.

If we take $|Z| = +\sqrt{1+\tan^2 3\pi/5}$, then

$$|Z| = +\sec 3/5\pi$$

which is a negative quantity. But $|Z|$ is not negative. Hence the correct answer should be $-\sec 3\pi/5$.

Hence to find the correct value of $|Z|$, we should use

$$|Z| = \sqrt{x^2+y^2}.$$

b) Again to find the real and imaginary parts of the complex number $3 + i\sqrt{-2}$, one may think that the real part is 3, but it is not. It should first be expressed in the form $a + ib$.

ie. $3 + i\sqrt{-2} = 3 + i\sqrt{2(-1)} = 3 + i\sqrt{2}\, i$ (Taking $i = \sqrt{-1}$)

$$= (3 - \sqrt{2}) + i.0$$

$\therefore$ Real part of $3 + i\sqrt{-2} = 3 - \sqrt{2}$

and Imaginary part $= 0$.

c) To find the argument θ of a complex number $z = x + iy$ if we consider only $\theta = \tan^{-1}(y/x)$, then we may have a wrong answer. [illegible] argument it is better to consider $\cos\theta = x/r$ and $\sin\theta = y/r$ simultaneously where r is the modulus of z. This will be clear from the following example;

(i) Find the principal value of the argument of

$- \quad -\sqrt{3}/2\, i$.

Let $z = -1/2 - \sqrt{3}/2\, i = x + iy$

$\therefore x = -\frac{1}{2}, \quad y = -\sqrt{3}/2$.

If we have $\theta = \tan^{-1}(y/x) = \tan^{-1}\sqrt{3} = \pi/3$,

$\therefore$ The principal value is $\pi/3$, which is incorrect.

Now $|z| = \sqrt{(-\frac{1}{2})^2 + (\sqrt{3}/2)^2} = 1$.

$\therefore \cos\theta$ [illegible] and [illegible] $\theta =$ [illegible]

$\therefore$ The argument is $(\pi + \pi/3) > \pi$

$\therefore$ The principal value of the argument is

$$-2\pi + (\pi + \pi/3) = -2\pi/3.$$

which is the correct principal value of the argument.

d) Some teachers may have the idea that $-i$ is negative of i. This is a conceptually wrong idea as i is neither positive nor negative.

6. **Gaps in the NCERT text book.**

(1) Although 'Laplace De[illegible] theorem' has been shown to be valid for integral indices and it has been written in the NCERT text book (Art 2.[illegible]) page [illegible], it is desirable that clear statement of this theorem may be given.

(2) Factorisation of $a^3 + b^3$, $a^3 - b^3$, and $a^3 + b^3 + c^3 - 3abc$ in linear factors as given below may be introduced

$$a^3 + b^3 = (a + b)(a + wb)(a + w^2 b)$$

$$a^3 - b^3 = (a - b)(a - wb)(a - w^2 b)$$

$$a^3 + b^3 + c^3 - 3abc = (a + b + c)(a + bw + cw^2)$$
$$(a + cw^2 + cw)$$

7. **Discussions of some interesting questions that may be asked by the teachers to the resource persons.**

(1) why $-\pi < \theta \le \pi$ is taken as the principal value of the argument ? What is the harm in takeing the principal value of the argument to be $-\pi \le \theta < \pi$

This can be treated as a convention. There is no harm in takeing the principal value of the argument to be $-\pi \le \theta < \pi$.

(2) If $z_1 = x_1 + iy_1$ and $z_2 = x_2 + iy_2$ such that $x_1 > x_2$, $y_1 > y_2$, then can we say that $z_1 > z_2$?

The answer is " No ".

The complex numbers are not ordered.

3. The argument of $x + iy$ is defined to be $\tan^{-1}(y/x)$, then what is the value of argument of $0 + i0$?

In fact it is undefined.

8. <u>Discussions of some enrichment materials in the topic which the teachers are supposed to know.</u>

(1) If we write $\sqrt{-16}\cdot\sqrt{-4} = \sqrt{(-16) \times (-4)} = \sqrt{64} = 8$, an error is committed. This is an incorrect result in applying the formula $\sqrt{a}\cdot\sqrt{b} = \sqrt{ab}$, which holds good only if a and b are both positive or one is negative. It does not hold good if a and b are both negative, i.e. if $\sqrt{a}$ and $\sqrt{b}$ are imaginary numbers. To avoid this difficulty, we should write the expressions of the form $\sqrt{-b}$, $b > 0$ in the form $i\sqrt{b}$ before any mathematical operation is that.

(2) Real number system is ordered. Complex numbers are not ordered, as in the system of real numbers. We can not say that one complex number is greater than or less than another complex number.

(3) The integral powers of i may have the values ± 1 or, $\pm i$ only.

$i^1 = i$	$i^{4m} = 1$	$i^{-1} = -i$	$i^{-4m} = 1$
$i^2 = -1$	$i^{4m+1} = i$	$i^{-2} = -1$	$i^{-(4m+1)} = -i$
$i^3 = -i$	$i^{4m+2} = -1$	$i^{-3} = i$	$i^{-(4m+2)} = -1$
$i^4 = 1$	$i^{4m+3} = -i$	$i^{-4} = 1$	$i^{-(4m+3)} = i$

where m is an integer

(4) The integral powers of w take the values $1, w, w^2$; $w^{3n} = 1$, $w^{3n+1} = w$, $w^{3n+2} = w^2$ where n is an integer.

9. <u>Construction of one (or more) fallacious [illegible] (if possible) to test a particular concept and its solution.</u>

1. Explain the fallacy of the following problem :

We have an identity

$$\sqrt{x - y} = \sqrt{-(y - x)} = i\sqrt{y - x} \quad \ldots\ldots \quad (1)$$

We put $x = a$, $y = b$, we get (where a and b are unequal and a and b are real numbers).

$$\sqrt{a - b} = i\sqrt{b - a} \quad \ldots\ldots \quad (2)$$

Again in (1) we put

$$\sqrt{b - a} = i\sqrt{a - b} \quad \ldots\ldots \quad (3)$$

Multiplying (2) and (3) together

$$\sqrt{a - b} \times \sqrt{b - a} = i^2\sqrt{a - b}\sqrt{b - a}$$

or, $1 = i^2$ (cancelling by $\sqrt{a - b}\sqrt{b - a}$ from both sides)

or, $1 = -1$.

<u>Answer</u> (1) When a, b are real numbers,

then either $a > b$ or $b > a$ $\therefore$ $\sqrt{a - b} = i\sqrt{b - a}$

and $\sqrt{b - a} = i\sqrt{a - b}$

cannot hold simultaneously.

Hence, the fallacy.

(2) When a and b are both complex numbers, in that case both (2) and (3) [illegible] simultaneously as for a complex number z, $z^{1/2}$ gives multiple values.
Hence the fallacy.

2. Prove that for any two complex numbers z_1, z_2.

$$Re\left(\frac{z_1}{z_1+z_2}\right) + Re\left(\frac{z_2}{z_1+z_2}\right) = 1$$

Solution :

$$Re\left(\frac{z_1}{z_1+z_2}\right) = \frac{1}{2}\left(\frac{z_1}{z_1+z_2} + \frac{\overline{z_1}}{\overline{z_1}+\overline{z_2}}\right)$$

$$Re\left(\frac{z_2}{z_1+z_2}\right) = \frac{1}{2}\left(\frac{z_2}{z_1+z_2} + \frac{\overline{z_2}}{\overline{z_1}+\overline{z_2}}\right)$$

$$\therefore \quad Re\left(\frac{z_1}{z_1+z_2}\right) + Re\left(\frac{z_2}{z_1+z_2}\right)$$

$$= \frac{1}{2}\left(\frac{z_1}{z_1+z_2} + \frac{\overline{z_1}}{\overline{z_1}+\overline{z_2}}\right) + \frac{1}{2}\left(\frac{z_2}{z_1+z_2} + \frac{\overline{z_2}}{\overline{z_1}+\overline{z_2}}\right)$$

$$= \frac{1}{2}\left(\frac{z_1+z_2}{z_1+z_2} + \frac{\overline{z_1}+\overline{z_2}}{\overline{z_1}+\overline{z_2}}\right)$$

$$= 1.$$

3. <u>Problem</u>: Prove that the triangle whose vertices are the points Z_1, Z_2, Z_3 on the Argand diagram is an equilateral triangle if and only if $Z_1^2 + Z_2^2 + Z_3^2 = Z_1 Z_2 + Z_2 Z_3 + Z_3 Z_1$

<u>Ans</u>: The lengths of the sides are $|Z_1 - Z_2|$, $|Z_2 - Z_3|$, $|Z_3 - Z_1|$

Since, the triangle is an equilateral triangle, $|Z_1 - Z_2| = |Z_2 - Z_3| = |Z_3 - Z_1|$ $\therefore |Z_1 - Z_2|^2 = |Z_2 - Z_3|^2 = |Z_3 - Z_1|^2$ .... (A)

Taking first two, $|Z_1 - Z_2|^2 = |Z_2 - Z_3|^2$

i.e. $(\bar{Z}_1 - \bar{Z}_2)(Z_1 - Z_2) = (\bar{Z}_2 - \bar{Z}_3)(Z_2 - Z_3)$

or, $\dfrac{Z_1 - Z_2}{\bar{Z}_2 - \bar{Z}_3} = \dfrac{Z_2 - Z_3}{\bar{Z}_1 - \bar{Z}_2} = \dfrac{Z_1 - Z_2 + Z_2 - Z_3}{\bar{Z}_2 - \bar{Z}_3 + \bar{Z}_1 - \bar{Z}_2}$

or, $\dfrac{Z_1 - Z_2}{\bar{Z}_2 - \bar{Z}_3} = \dfrac{Z_1 - Z_3}{\bar{Z}_1 - \bar{Z}_3}$ ...... (1)

Also we have (from (A)) $|Z_2 - Z_3|^2 = |Z_3 - Z_1|^2$

$\therefore (Z_2 - Z_3)(\bar{Z}_2 - \bar{Z}_3) = (Z_3 - Z_1)(\bar{Z}_3 - \bar{Z}_1)$ ...... (2)

Multiplying (1) and (2) together, we get

$$(Z_1 - Z_2)(Z_2 - Z_3) = (Z_3 - Z_1)^2$$

or, $Z_1 Z_2 - Z_2^2 - Z_1 Z_3 + Z_2 Z_3 = Z_3^2 + Z_1^2 - 2Z_3 Z_1$

or, $Z_1^2 + Z_2^2 + Z_3^2 = Z_1 Z_2 + Z_2 Z_3 + Z_3 Z_1$.

<u>Conversely</u>, given $Z_1^2 + Z_2^2 + Z_3^2 = Z_1 Z_2 + Z_2 Z_3 + Z_3 Z_1$.

or, $Z_1 Z_2 - Z_2^2 + Z_2 Z_3 = Z_3^2 + Z_1^2 - Z_3 Z_1$

or, $z_1 z_2 - z_2^2 - z_3 z_1 + z_2 z_3 = z_3^2 + z_1^2 - 2z_3 z_1$.

or, $(z_1 - z_2)(z_2 - z_3) = (z_3 - z_1)^2$ ...... (3)

$\therefore$ $(\bar{z}_1 - \bar{z}_2)(\bar{z}_2 - \bar{z}_3) = (\bar{z}_3 - \bar{z}_1)^2$ .... (4)

[If $w = z_1 z_2$ then $\bar{w} = \overline{z_1 z_2}$ i.e. $\bar{z}_1 . \bar{z}_2 = \bar{w}$.]

Multiplying (3) and (4) together, we get

$$(z_1 - z_2)(\bar{z}_1 - \bar{z}_2)(z_2 - z_3)(\bar{z}_2 - \bar{z}_3) = (z_3 - z_1)^2 (\bar{z}_3 -$$

i.e. $|z_1 - z_2|^2 \, |z_2 - z_3|^2 = |z_3 - z_1|^4$

$\therefore$ $|z_1 - z_2| \, |z_2 - z_3| = |z_3 - z_1|^2$ ..... (5)

Similarly, $|z_2 - z_3| \, |z_3 - z_1| = |z_1 - z_2|^2$ .... (6)

and $|z_3 - z_1| \, |z_1 - z_2| = |z_2 - z_3|^2$ .... (7)

From (7) $|z_1 - z_2| = \dfrac{|z_2 - z_3|^2}{|z_3 - z_1|}$

Substituting in (6), $|z_2 - z_3| \, |z_3 - z_1| = |z_2 - z_3|^4 / |z_3 - z_1|$

or, $|z_3 - z_1|^3 = |z_2 - z_3|^3$

$\Longrightarrow$ $|z_3 - z_1| = |z_2 - z_3|$.

Similarly, it can be shown from (5) and (6) that

$$|z_2 - z_3| = |z_1 - z_2|.$$

Hence $|z_1 - z_2| = |z_2 - z_3| = |z_3 - z_1|$

i.e. the triangle with vertices z_1, z_2, z_3 is an equilateral triangle.

10. References (of easy, cheap and easily available books)

The following books can be referred to teachers :

1. College Algebra and Trigonometry - by J.S. Ratti (MacMillan Publishing Co., Inc. New York)

2. Higher Algebbra - by A. Kurosh (Mir Publishers)

3. Higher Algebra - by Ghosh and Chakraborty (U.N. Dhur & Co., Calcutta)

4. Higher Algebra - by Maity and Ghosh (Central Book Agency, Calcutta)

— X —

SETS, RELATIONS AND FUNCTIONS

Prepared by

1. Dr S. R. Joshi
Department of Mathematics
Yogeshwari Mahavidyalaya
Ambagogai, Distt. Beed
Maharashtra - 431517

2. Dr (Mrs) Indira Datta
Lady Brabourne College
Department of Mathematics
P 1/2, Suhrawardy Avenue
Calcutta - 700017

SETS, RELATIONS AND FUNCTIONS.

1. Motivation of the topic:

The concept of set may be motivated by considering daily life situations. Many times we talk about our family, club, political parties, organizations, institutions and so on. All these are examples of sets from mathematical point of view. A person can be a member of a political party and also a member of an organization. All these examples motivate not only the concept of a set but also the concept of membership, intersection of two sets, disjoint sets etc.

Sometimes we talk about a set of books, bunch of papers, tea set etc. In such cases the meaning of set is somewhat different from the usual meaning of set we consider from mathematical point of view.

The concept of 'Relation' may be motivated by considering the usual relations in each family. If x and y are two members of a family then x can be a sister of y or y can be a daughter of x and so on. If there are, say, 200 families in a village, we may define the relationship between two persons if they belong to a unique family. This relation in fact is an equivalence relation and the equivalence classes are the families of that village.

The idea of a function may be motivated by considering a set of colours and a set of birds. Every bird has a unique colour. Thus if x is a bird then the colour of x may be denoted by f(x) or c(x) or any other convenient symbol.

The concept of binary operation on a set may be motivated by considering the usual operations of addition, multiplication etc. The following two examples will also be useful to understand the concept.

<u>Ex. 1</u> : Let U be a universal set so that the other sets are all subsets of U. Given any two sets A and B we obtain a third set say $A \cap B$. Then intersection, $\cap$ is a binary operation on U. Since $\cap$ is a function from U X U to U.

<u>Ex. 2</u>: Consider the set S of all statements. We know that every statement is either true or false but not both. If p and q are two statements, then p or q is also a statement. Thus 'or' is a binary operation on S. It is generally denoted by $\vee$.

2. <u>Brief Outline of the content</u>

In the book, algebraic operation on sets are defined and some results concerning them, like commutative and associative properties of union and intersection of sets, De Morgan's laws etc are mentioned.

Next, the definitions of a binary relation R from a non-empty set A to another non-empty set B' and a binary relation R on a non-empty set A' have been given. The definition of equivalence relation on a set comes thereafter.

The idea of functions or mappings comes next as a particular case of a relation. Here the difference between a relation $R : A \longrightarrow B$ and a function $f : A \longrightarrow B$ has been made clear.

Next, the idea of binary operations on a set has been dealt with.

3. Explanation of technical/mathematical terms not properly explained in the text books.

The definition of binary operation is given on page 13 of the book as "a map: $A \times A \longrightarrow B$ is called a binary operation in A and if $B \subset A$, then it is said to be closed with reference to the operation."

But, this definition is confusing and we think that the definition of binary operation on a non-empty set A can be given as a mapping from $A \times A \longrightarrow A$.

4. Alternative easier approach, if any, in discussing some subtopics.

On page 9 of the book, the definition of injective map $f : A \longrightarrow B$ is given as follows :

If $x_1, x_2 \in A$, and $x_1 \neq x_2$ implies $f(x_1) \neq f(x_2)$, then f is said to be an injective function or one-one function.

Here we like to mention that in solving problems where injectivity of a function is to be tested, it is more useful if we follow the definition as :
A function $f : A \rightarrow B$ is injective, if $x_1, x_2 \in A$, then

$$f(x_1) = f(x_2) \Rightarrow x_1 = x_2$$

For example $f(x) = 3x - 7$ is injective, since

$$f(x_1) = f(x_2) \Rightarrow 3x_1 - 7 = 3x_2 - 7$$
$$\Rightarrow 3x_1 = 3x_2$$
$$\Rightarrow x_1 = x_2.$$

Note that in order [illegible] that a certain map is not injective or not one-one, then it is sufficient to give one counter example:

e.g. $f(x) = x^{[illegible]}$ is not one-one.

Because $f(\ ) = f(\ )$ and [illegible] = [illegible]

5. **<u>Basic concepts to be explained in teaching the topic:</u>**

The various concepts about sets, relations, functions and binary operations which need emphasis are listed below in brief:

(i) A set is an undefined term and it can be described with the help of some synonymous words like collection, family, etc.

(ii) A set may be said to be a subset of itself since this follows from the definition of subset.

(iii) A universal set [illegible] that it is the set of all things in the universe. A universal set need not be unique. There can be different [illegible]. For example, when we want to study plane geometry, the universal set is the plane itself.

(iv) The concept of "complement of a set" can be thought only when we have some universal set at our disposal.

(v) $A - B = B^c$ only when A is the universal set.

(vi) Given a relation R in A, it is not necessary that it should be either reflexive, symmetric or transitive. This follows from the definition of a relation that it can be any subset of $A \times A$.

(vii) Given a set A we can have different equivalence relations on A.

(viii) A function $f : A \longrightarrow B$ is a particular kind of relation from A to B and that if $f(a) = b$ and $f(a) = c$, then $b = c$. This is the same as saying that

$$a \; R \; b \text{ and } a \; R \; c, \text{ then } b = c$$

(ix) A binary operation on A is a function from $A \times A$ to A. The image of (a, b) under a binary operation say f, is generally denoted by $a \circ b$, $a + b$ or $a \cdot b$ etc. . . . The usual basic operations on numbers i.e. addition, multiplication etc. are all familiar examples of binary operations.

(x) A binary operation need not be always commutative, i.e. given a and b in A and + is a binary operation, it need not be always true that $a + b = b + a$. e.g. in the set of vectors the operation of cross product is not commutative i.e. $\vec{A} \times \vec{B} \neq \vec{B} \times \vec{A}$.

6. <u>Misconceptions about some points and certain gaps.</u>

Some teachers think that there is only one universal set as the name itself suggests. But this is not the case. There can be different universal sets according to different situations.

Some teachers feel that a set A can not be a subset of itself. This happens so because they think that $A \subseteq B$ means that A is a proper subset of B and that $A \neq B$. Infact, in the definition of $A \subseteq B$, the case $A = B$ is also taken into account.

Neither the set $\{0\}$ nor the set $\{\phi\}$ is a null set. The first is a set consisting of a single element zero while the second is a set consisting of one element ϕ which is the null set. A member of a set can be a set also. In such a case it is called a family.

Note that a set can not be a member of itself i.e. $A \in A$ is meaningless. Because of this a set of all sets does not exist. Because if such a set, say S, exists then $S \in S$ and this is impossible.

Given a relation R in A and if $a, b \in A$, then it is not necessary that a R b or b R a holds, e.g. consider the relation of divisibility on the set N of all positive integers, then 3 R 8 does not hold since 3 does not divide 8.

It is not necessary that for a given set A there exists only one equivalence relation, e.g. in the set N of positive integers both of the following relations are equivalence relations.

(i) a R b means [illegible] divides $(a - b)$

(ii) a R b means $a = b$

It should be noted that as far as the second relation is concerned, there are as many equivalence classes as there are natural numbers.

Some teachers feel that for a given function f if $f(a) = f(b)$, then $a = b$. But this is not the case always. Infact this happens only when f is an injective map. For example, for the function

$f : R \longrightarrow R$ given by $f(x) = x^2$,

we see that $f(4) = f(-4) = 16$

But $4 \neq$ [illegible].

Many teachers don't appreciate the point that a binary operation is a kind of function. This happens because while dealing with binary operations we don't use the usual symbols such as f, g, f_1, F etc. Instead, we use symbols such as $+$, $-$, $\cdot$ $\times$ etc.

For example consider addition of two numbers in R. We use $a + b$ to denote the image of (a, b). In fact the function

$+ : R \times R \longrightarrow R$ it self is known as addition. Further $+(a, b)$ is written as $a + b$.

Every binary operation need not be commutative. For example, the operation of matrix product in the set of 2×2 matrices is not commutative.

Another practical example is the following :

The father of a brother $\neq$ a brother of the father.

On page 16 of the NCERT book in problem 5, the word partition is mentioned, assuming that the concerned teachers already know the definition of partition. Unless one has a clear cut idea about partition, one can not solve the problems regarding partition. For the sake of completeness, we have given the definition of partition in section 8 of this topic along with other related material.

7. <u>Some Interesting Questions for Thorough Understanding:</u>

The following questions may be useful for teachers for their thorough understanding of different points in the chapter.

(i) Does there exist a set A such that $A = A^c$? If so, what is the universal set ?

(ii) Can we define a set ? If not why ? Are there any terms in mathematics which are undefined ? If yes what are they ?

(iii) If $A \cup B = A$ and $A \cap B = B$, what is the relation between A and B ?

(iv) If a set A is divided into 3 disjoint subsets, is it possible to define an equivalence relation in A ? Justify your answer.

(v) What is the usual and familiar relation in the family of sets? Is it symmetric ?

(vi) How many relations can we define in a set $A = \{a, b, c\}$? How many are equivalence relations ?

(vii) Prove that every function $f : A \rightarrow A$ is a relation. Is the converse true ? Justify your answer.

We shall now discuss in brief the solutions of the above-mentioned questions.

(i) $A = A^c$ holds true only when A is a null set and the universal set is also the null set.

(ii) Set is an undefined term since in the definition of a set, we use synonimous words. In Geometry, point, plane and line are undefined terms.

(iii) $B \subset A$.

(iv) If $A = A_1 \cup A_2 \cup A_3$, define a R b to mean that a & b belong to A_i for some $i = 1, 2, 3$. With respect to this relation R (R is an equivalence relation) A_1, A_2 and A_3 are equivalence classes.

(v) A R B means A is a subset of B. R is not symmetric.

(vi) There are 9 order pairs in A x A. Hence there are $2^9 - 1$ nonempty subsets of A x A. Hence there are 511 relations which can be defined in A. There are 5 equivalence relations that can be defined on A. Two of the five are as follows:

$$R_1 = \{(a, a), (b, b), (c, c)\}$$
$$R_2 = \{(a, a), (b, b), (c, c), (b, a), (c, b)\}$$

(vii) According to the definition of a function, every function is a relation. The converse is not true. The following example justifies the claim.

Let $A = \{a, b, c\}$
Let $f = \{(a, a), (b, c), (c, a), (b, a)\}$

Here f is a relation since $f \subset A \times A$ but f is not a function because b has two images under f i.e. $f(b) = c$ and $f(b) = a$ which is impossible for a function.

(viii) In order to divide N into 3 disjoint subsets, it is not necessary to use the concept of equivalence relation. But in every such division the idea of equivalence relation is involved.

Consider the relation R defined by

a R b means 3 divides a - b.

Then R is an equivalence relation. There are only 3 equivalence classes and they are

$$A_1 = \bar{1} = \{1, 4, 7, 10, 13, \ldots\}$$
$$A_2 = \bar{2} = \{2, 5, 8, 11, 14, \ldots\}$$
$$A_3 = \bar{3} = \{3, 6, 9, 12, 15, \ldots\}$$

(ix) Consider the set of vectors in a three dimensional space and the operation as the cross product of two vectors. Then this binary operation is not associative. i.e.

$A \times (B \times C) = (A \times B) \times C$ does not hold always. e.g.

$$\left[(\hat{i} + \hat{j}) \times \hat{k}\right] \times (\hat{i} + \hat{k}) \neq \left[(\hat{i} + \hat{j}) \times \left\{\hat{k} \times (\hat{i} + \hat{k})\right\}\right]$$

8. <u>Discussion on some enrichment materials on the topic:</u>

A teacher of Mathematics is already familiar with certain kinds of relations on a set. There are two important types of relations on a set (i) Equivalence relation and (ii) Partial order relation.

The concept of an <u>equivalence relation</u> is an extremely important one, and plays a central role in all of mathematics. The definition of equivalence relation can be given precisely as follows :

Let A be a given non-empty set.

A subset R of $A \times A$ is said to be an equivalence relation on A, if

i) $(a, a) \in R$ for all $a \in A$

ii) $(a, b) \in R$ implies $(b, a) \in R$ for any two element $a, b \in A$

iii) $(a, b) \in R$ and $(b, c) \in R$ imply that $(a, c) \in R$, for any three elements $a, b, c \in A$.

These three properties are respectively referred to as reflexivity, symmetry and transitivity.

Whenever an equivalence relation is defined on a nonempty set A, there arises the concept of equivalence class of an element $a \in A$. We define it as follows :

Let A be a nonempty set and let R be an equivalence relation on the set A. Then the equivalence class of $a \in A$ is the set $\{x \in A \mid (a, x) \in R\}$. We denote this set by $Cl(a)$ or $[a]$. Thus $[a]$ or $cl(a)$ is the set of all those elements of A which are equivalent to a.

We now state the most fundamental theorem in this connection.

The distinct equivalence classes of an equivalence relation on A provides us with a decomposition of A as a union of mutually disjoint subsets. Conversely, given a decomposition of A as a union of mutually disjoint, nonempty subsets, we can define an equivalence relation on A for which these subsets are the distinct equivalence classes.

Proof: Let A be a nonempty set and R be an equivalence relation on A. Then, for every $a \in A$, $(a, a) \in R \Rightarrow a \in cl(a)$. Hence the union of the $cl(a)$'s [illegible] A. Next, we assert that given two equivalence classes, say, $cl(a)$ and $cl(b)$, they are either equal or disjoint.

Suppose that $cl(a)$ and $cl(b)$ are not disjoint, i.e. there is an element, say x, $\in cl(a) \cap cl(b)$. We shall prove that $cl(a) = cl(b)$ in this case.

$\because \quad x \in cl(a)$

$\quad (a, x) \in R$,

again $\quad x \in cl(b)$ also.

$\therefore \quad (b, x) \in R$

$\quad$ Now, $(b, x) \in R \Rightarrow (x, b) \in R$ as R is symmetric

$(a, x) \in R$ and $(x, b) \in R$

$\Rightarrow \underline{(a, b) \in R}$ $\qquad \because$ R is transitive

Let y be any element of $cl(b)$. Then $(b, y) \in R$

$\therefore \quad (a, b) \in R$ and $(b, y) \in R$

$\Rightarrow (a, y) \in R$

i.e. $\quad y \in cl(a)$

Hence $\quad cl(b) \subset cl(a)$

Similarly, we can prove $cl(a) \subset cl(b)$

$\therefore \quad \underline{cl(a) = cl(b)}$

We have thus shown that the distinct cl(a)'s are mutually disjoint and that their union is A. Hence, the first part of the theorem is proved. Conversely, suppose that

$$A = \bigcup A_i \quad \text{where } A_i \text{ are mutually disjoint non-empty sets.}$$

Now, for any $a \in A$, we can say that a is in exactly one A_i of the family $\{A_i\}$. We now define a binary relation R for $a, b \in A$ as $(a, b) \in R$ if a, b are in the same A_i. Then it can be easily verified that this binary relation R is an equivalence relation.

The fundamental theorem about equivalence relation can be stated in a simpler way with the introduction of the term partition which we define as follows :

Let $\{B_i\}_{i \in I}$ where I is the index set be a family of non-empty subsets of a set A. Then $\{B_i\}_{i \in I}$ is called a partition of A if (i) $\bigcup_{i \in I} B_i = A$

and (ii) for any subsets B_i and B_j, either $B_i = B_j$ or $B_i \cap B_j = \phi$, [ϕ is the null set]

Then alternative statement of the previous theorem can be given as :

An equivalence relation R on a set A determines a partition of A. Conversely, each partition of A yields an equivalence relation on A.

Next, we deal in detail with the other type of relation known as partial order relation.

A relation R in a set A is called a partial order relation (or only an order relation) if

(a) a R a holds for all $a \in A$

(b) a R b and b R a implies $a = b$ for $a, b \in A$.

(c) if a R b and b R c, then a R c where $a, b, c \in A$.

We say a relation R in A is antisymmetric if the property (b) is satisfied.

Thus a relation R in A is an order relation if it is reflexive, antisymmetric and transitive.

<u>Notation</u>: It is a general practice to express the relation a R b by writing $a \leq b$. Note that $a \leq b$ does not always mean that a is less than or equal to b. $a \leq b$ means some times a divides b or a is not equal to b or a is a subset of b.

The usual examples of partial order relation are as follows:

(i) In the set of real numbers R_1, we define $a \leq b$ to mean that a is less than or equal to b. $\leq$ is a partial order relation.

(ii) Let Z be the set of integers. Then by $a \leq b$ we mean $a \neq 0$ and a divides b. We generally write a/b (read as a divides b or b is divisible by a) for $a \leq b$.

(iii) In the family of sets U we have the usual relation $A \subset B$ where $A, B \in U$. It can be verified that $\subset$ is a partial order relation in U.

(iv) Let R^n be the set of n-tuples of real numbers. Let

$$x = (a_1, a_2, \ldots, a_n)$$

and

$$y = (b_1, b_2, \ldots, b_n)$$

Then by $x \leq y$ we mean that $a_i \leq b_i$ for all $i = 1, 2, \ldots n$. Where by $a_i \leq b_i$ we mean that a_i is less than or equal to b_i. It can be verified that the relation defined above is a partial order relation.

A set P equipped with a partial order relation $\leq$ is called a partial order set or simply a POSET.

If S is a subset of a poset $(P, \leq)$, then an element $a \in P$ is called an upper bound of S if $x \leq a \ \forall \ x \in S$.

Note that we also write $b \geq a$ to indicate $a \leq b$ ($b \geq a$ may be read as b is greater than or equal to a).

Similarly, an element $b \in P$ is called a lower bound of S if $b \leq x \ \forall \ x \in S$.

Note that an upper bound or a lower bound of a set S need not be unique. Further an upper bound or lower bound of a set may not exist in certain cases. For instance, if P = set of all positive integers and E is the set of even positive integers, then $E \subset P$. Define $a \leq b$ to mean that a divides b. Then E has no upper bound and that 1 is the only lower bound of E.

If a happens to be the smallest among all upper bounds of a set S then a is called the least upper bound of S.

Similarly, the greatest among all the lower bounds of S is called the greatest lower bound of S. It can be noted that if the least upper bound (lub) or greatest lower bound (glb) of a set S exists, then it is unique.

For example, consider the partial order relation of divisibility in the set N of all natural numbers.

Let $S = \{$ [illegible], 5, 7, 6 $\}$

Then it can be observed by inspection that 210 is the lub of S since $x \mid 210 \ \forall x \in S$. Note 210 is the least common multiple of the integers in S. Note that $210 \notin S$. Similarly 1 is the glb of S.

If U_1 is any subfamily of a given family U of sets. Then the null set is the glb of U_1 and the union of all sets in U_1 is the lub of U_1.

An element $x \in S$, where S is a subset of a poset $(P, \leq)$ is said to be a maximal element of S if there does not exist $y \in S$ satisfying $x \leq y$.

Similarly $x \in S$ is called a minimal element of S if there does not exist any element $y \in S$ satisfying $y \leq x$. The following two examples will illustrate the concept of maximal and minimal elements of a set.

<u>Ex.1</u>: Let us consider the divisibility relation in N.

Let $S = \{ 3, 5, 6, 10, 14 \}$.

Then 10 and 14 are maximal elements of S since both of them satisfy the condition of maximality. Similarly 3, 5 and 14 are minimal elements of S.

Ex. 2: Consider the relation $\subseteq$ (is a subset of) in the family U of sets given by

$$U = \{\{a\}, \{b\}, \{c\}, \{a, b\}, \{a, c\}, \{b, c\}, \{a, b, c\}\}$$

$$U_1 = \{\{a\}, \{b\}, \{a, c\}, \{b, c\}\}$$

Here $A = \{a, c\}$ & $B = \{b, c\}$ are maximal elements and $\{a\}, \{b\}$ are minimal elements.

The concept of maximal element is very important in many mathematical situations. Related to this there is one important axiom known as AXIOM OF CHOICE or also known as Zorn's Lemma.

Before stating the Axiom of Choice we first give the definition of a chain.

A subset S in a POSET $(P, \leq)$ is said to be a chain if $a \leq b$ holds or $b \leq a$ holds for every $a, b \in S$.

e.g. $S = \{3, 6, 12, 24, 120, 360\}$ is a chain in N with respect to the relation of divisibility.

Similarly, $\{\{a\}, \{a, b\}, \{a, b, d\}\}$ is a chain of 3 elements in a family of sets with respect to the relation of "subset of".

Axiom of choice: If every chain in S has the least upper bound in S, then maximal element of S exists.

This result is used in some branches of Mathematics. For instance, we apply this in proving certain theorems of Topology, Linear Algebra, Functional Analysis and other branches of mathematics.

9. Some Interesting Problems

(i) Let $A = \{3, 5, 6, 10\}$

In A define a relation a R b to mean that a divides b. We say that $x \in A$ is a maximal element of A if there does not exist any $y \in A$ such that x R y. Find all the maximal elements of A.

Solution : We observe that

3 R 6, 5 R 10. Also 6 R 10 does not hold. Hence 6 and 10 are both maximal elements.

(ii) In a family of sets, define two binary operations . and $\triangle$ as follows :

(a) $A . B = A \cap B$

(b) $A \triangle B = (A \cup B) - (A \cap B)$

or $A \triangle B = (A - B) \cup (B - A)$

Then prove that

$$A . (B \triangle C) = A.B \triangle A.C$$

Hint: Mathematical proof of this problem is some what difficult. But with the help of Venn Diagram, it becomes easier to get a solution. It is known that a family of sets equipped with the two binary operations $\cdot$ and $\triangle$ is a ring and this concept of ring is used in the study of Measure Theory.

(iii) The property (ii) of an equivalence relation states that $(a, b) \in R \Rightarrow (b, a) \in R$; property (iii) states that $(a, b) \in R$ and $(b, c) \in R \Rightarrow (a, c) \in R$. What is wrong with the following proof that properties (ii) and (iii) imply property (i) ?

Let $(a, b) \in R$; then $(b, a) \in R$ [by Prop (ii)] whence by property (iii) $\underline{(a, a) \in R}$.

Solution: In the proof it is assumed that $(a, b) \in R$ holds without mentioning any restriction on a and b. In other words, it is assumed that a & b are related for any two elements $a, b \in A$. This basic assumption leads that $(a, a) \in R$. Hence the fallacy lies in the assumption that $(a, b) \in R$ holds for every pair (a, b).

10. References :

The following books may be useful for teachers and resource persons :

(i) Finite Mathematics by Kemeny Thompson and Snell.

(ii) Set Theory and Logic by Robert Stoll.

(iii) Topics in Algebra by I.N. Herstein published by Vani Educational Books.

(iv) Modern Algebra by Leadership Project, Bombay University, Bombay.

(v) Modern Algebra by Neal H. McCoy.

(vi) Schaum's Outline of Theory and Problems of Set Theory and Related Topics – Seymour Lipschutz published by : McGraw Hill International Book Company.

(vii) Higher Algebra : [illegible]
Published by Asoke Prakasan, Calcutta.

VECTORS AND THREE-DIMENSIONAL GEOMETRY

Prepared by

1. Dr A. K. Pal
   Department of Mathematics
   Jadavpur University
   Calcutta - 780032

2. Shri Harihar Ghosh
   Department of Mathematics
   Presidency College
   College Street
   Calcutta - 700073

1. **Motivation**

In nature, there are some quantities which can be completely defined by a single number. These quantities are called scalars. In contrast to these, there are quantities which are not fully defined by such single number but need something more. For instance, let us consider displacement of a particle from a point A to a distance of 5 Cms. When the point A remains fixed in space. This displacement through a distance of 5 Cms may occur in an infinitely many ways. All such points specifying displacements of a distance of 5 Cms from A in an arbitrary manner lie on a sphere of radius 5 Cms with the point A as centre. Hence to specify this displacement what we need in addition to the length 5 Cms, is the direction of such displacement. We thus find that the displacement is an example of quantities which need both magnitudes and direction for their complete definition. We call such quantities as Vectors. Displacements, velocities, accelerations, forces are vector quantities while temperature, mass etc. are scalar quantities.

For the definition of a vector quantity, we take the help of displacement as an example. We shall develop different aspects of vectors with the help of the idea of displacement of a particle.

2. **Brief Outline of the content**

The definition of a vector, different types of vectors, algebra of vectors etc. have been discussed in the textbook. The dot product and cross products of two vectors have also been discussed. The scalar triple product and the vector triple product of three vectors are also included. The application of vectors in finding the equation of a plane, straight line and sphere and the related topics in three dimensional geometry has also been included in the textbook.

3. **Explanation of technical/mathematical terms not properly explained in the textbook:**

Zero or Null Vector

In the book submitted by the NCERT, the term zero or the Null vector has not been used by the authors. They have used the phrase identity element for vector addition to represent a zero vector. The resource person should know that these two are the same. [illegible] 75, it is wrongly written that an identity element [illegible] addition (Null or zero vector) [illegible] direction.

An entity having no direction may mean that it is a scalar quantity. It is a vector whose magnitude is zero and can have any direction.

Negative Vector

If the sum of two vectors is a null vector then each of the two vectors is the negative of the other.

Unit Vector

A vector having unit magnitude is called a unit vector. If a vector is divided by it's modulus, we find a unit vector.

Free Vector and localized Vector

Free vectors have no restriction regarding their initial or terminal points, while a localized vector occupies a definite position in space. Unless otherwise stated, we mean a free vector when we use the term vector.

Coinitial Vector

Vectors originating from the same point or vectors having the same base point are called Coinitial Vectors.

In coordinated geometry, it does not matter which frame is used — the right handed or the left handed frame. But in vectors, it matters much. The correct definition of right handed frame is as follows: Let a right handed screw (a screw that may be driven by the right hand) be placed along O so that the pointed end (the tip) points in the positive direction of the Z-axis. If a rotation of the screw in the sense that carried Ox to Oy in a rotation of 1 right $\angle$ makes the screw advance along OZ, then the frame is called a right handed frame.

In Page 473 of the textbook, Notes 2 and 3 say that the section formula is not defined for $m = -n$. But it is very important to observe that, when R is an between P and Q, $m + n$ is positive. When R is on PQ produced or on QP produced, PR and RQ are clearly different in magnitude. So $m : n$ cannot be $-1 : 1$ and consequently $m + n \neq 0$. Thus wherever R may be, $m + n \neq 0$. So the section formula is defined for all positions of R without any restriction. [The restriction imposed in the book is exactly similar to the restriction in this statement : $6 \div 3 = 2$, provided 3 does not vanish !]

Since the denominator never vanishes, a convenient section formula can be obtained by taking the ratio as $1-\lambda : \lambda$ for all λ.

4. Alternative approach, if any, in discussing some subtopics

Vector triple Product

$$\vec{a} \times (\vec{b} \times \vec{c}) = (\vec{a} \cdot \vec{c})\vec{b} - (\vec{a} \cdot \vec{b})\vec{c}$$

An alternative proof :

Let us suppose the three vectors $\vec{a}$, $\vec{b}$ and $\vec{c}$ be the following:

$\vec{a} = a_1\hat{i} + a_2\hat{j} + a_3\hat{k}$, $\vec{b} = b_1\hat{i} + b_2\hat{j} + b_3\hat{k}$, $\vec{c} = c_1\hat{i} + c_2\hat{j} + c_3\hat{k}$

$$(\vec{b} \times \vec{c}) = \begin{vmatrix} \hat{i} & \hat{j} & \hat{k} \\ b_1 & b_2 & b_3 \\ c_1 & c_2 & c_3 \end{vmatrix} = \hat{i}(b_2c_3 - c_2b_3) + \hat{j}(b_3c_1 - b_1c_3) + \hat{k}(b_1c_2 - c_1b_2)$$

$$\therefore \vec{a} \times (\vec{b} \times \vec{c}) = \begin{vmatrix} \hat{i} & \hat{j} & \hat{k} \\ a_1 & a_2 & a_3 \\ (b_2c_3 - c_2b_3) & (b_3c_1 - b_1c_3) & (b_1c_2 - c_1b_2) \end{vmatrix}$$

$$= \hat{i}\left\{a_2(b_1c_2 - c_1b_2) - a_3(b_3c_1 - b_1c_3)\right\} + \hat{j}\left\{a_3(b_2c_3 - c_2b_3) - a_1(b_1c_2 - c_1b_2)\right\}$$
$$+ \hat{k}\left\{a_1(b_3c_1 - b_1c_3) - a_2(b_2c_3 - c_2b_3)\right\}$$
$$= \hat{i}\left[b_1(a_2c_2 + a_3c_3) - c_1(a_2b_2 + a_3b_3)\right] + \hat{j}\left[b_2(a_3c_3 + a_1c_1) - c_2(a_3b_3 + a_1b_1)\right]$$
$$+ \hat{k}\left[b_3(a_1c_1 + a_2c_2) - c_3(a_1b_1 + a_2b_2)\right]$$
$$= \hat{i}\left[b_1(a_1c_1 + a_2c_2 + a_3c_3) - c_1(a_1b_1 + a_2b_2 + a_3b_3)\right] + \hat{j}\left[b_2(a_1c_1 + a_2c_2 + a_3c_3)\right.$$
$$\left. - c_2(a_1b_1 + a_2b_2 + a_3b_3)\right] + \hat{k}\left[b_3(a_1c_1 + a_2c_2 + a_3c_3) - c_3(a_1b_1 + a_2b_2 + a_3b_3)\right]$$
$$= (a_1c_1 + a_2c_2 + a_3c_3)(\hat{i}\,b_1 + \hat{j}\,b_2 + \hat{k}\,b_3) - (a_1b_1 + a_2b_2 + a_3b_3)(\hat{i}\,c_1 + \hat{j}\,c_2$$
$$+ \hat{k}\,c_3) = (\vec{a} \cdot \vec{c})\vec{b} - (\vec{a} \cdot \vec{b})\vec{c}$$

This proof is more general and straight-forward than the proof given in the book of NCERT.

5. Basic Concepts to be emphasized in teaching the topic

The various concepts that should be emphasized to the teachers engaged in teaching vectors and three dimensional geometry through vectors are the following:

i) Vectors and scalars and their distinctions.

ii) Concept of null vector.

iii) Triangle law of vectors.

iv) Coplanarity of vectors.

v) Different types of vector products.

 a) Dot product or scalar product of two vectors and their commutative property.

 b) Cross product or vector product of two vectors and their non-commutative nature.

 c) Triple product of vectors

$$\vec{a} \cdot (\vec{b} \times \vec{c}) = \text{Scalar}$$

$$\vec{a} \times (\vec{b} \times \vec{c}) = \text{vector}$$

vi) Coordinates of a point

vii) Distance between two points

viii) Direction ratios and Direction Cosines

6. Analysis of conceptual errors that may be committed by teachers in teaching the topic (in this context, mention gaps and misconception, if any, in the textbook) which may misguide the teachers and students

Equality of two vectors and Null vector

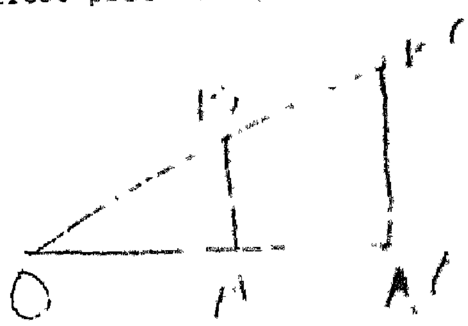
Two vectors are called equal only when they are equal in magnitude and possess the same direction. But in case of Null vectors, this may not be true. Null vector possesses some unique properties which are not found in case of other vectors.

Callinear and Parallel Vectors

Two Vectors are called parallel when they have the same direction but their lines of action are different whereas in the case of collinear vectors, their lines of action and direction both are identical.

Correct proof for the distributive law for multiplication of Vectors by real number

The proof for the distributive law for multiplication of vectors by real numbers given in the book of NCERT is confusing and not correct. The correct proof is given below:



Let $\overrightarrow{OA} = \vec{a}$ and $\overrightarrow{AB} = \vec{b}$

Let us suppose 'm' be any positive real number. Here in the adjoining figure $m\vec{a} = m\overrightarrow{OA} = \overrightarrow{OA'}$ where the directions of $\overrightarrow{OA}$ and $\overrightarrow{OA'}$ are the same but magnitude of $\overrightarrow{OA'}$ is m times that of $\overrightarrow{OA}$.

Let A' B' be drawn parallel to AB. We join O, B and extend it to meet A' B' at B' (say). Since $\triangle$s OA'B' and $\triangle$ OAB are similar, we find

$$\frac{OB'}{OB} = \frac{A'B'}{AB} = \frac{OA'}{OA} = m$$

Hence $\overrightarrow{A'B'} = m\overrightarrow{AB}$, since $A'B'$ and AB are parallel and length of the side $A'B'$ = m times that of AB.

So $\overrightarrow{OA'} + \overrightarrow{A'B'} = \overrightarrow{OB'}$

$m\overrightarrow{OA} + m\overrightarrow{AB} = m\overrightarrow{OB} = m(\overrightarrow{OA} + \overrightarrow{AB})$

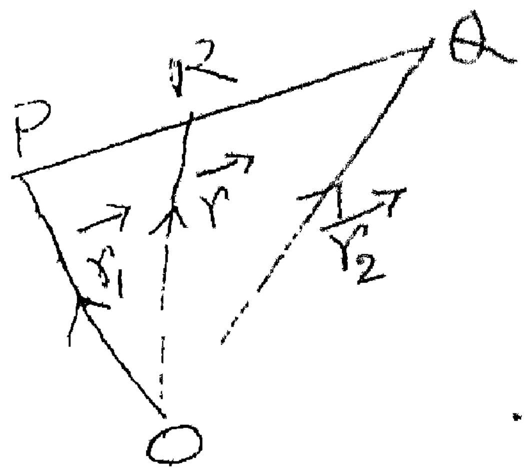
or $m\vec{a} + m\vec{b} = m(\vec{a} + \vec{b})$

The above relation is valid even when m is a negative scalar.

Section Formula

The coordinates of some point R which divides the line PQ in a certain ratio m : n has been derived twice in the chapter on Algebra of Vectors, Page 415 and in three-dimensional geometry, Page 472. The proof given in Page 415 is very round about and contains several unnecessary steps. So this proof may be deleted.

Condition of Collinearity of three points :



If the position Vectors of the points P and Q be $\vec{r_1}$ and $\vec{r_2}$ referred to O as origin, then the position Vector of any point R which divides the segment PQ in the ratio m : n has been derived in the form

$$\vec{r} = \frac{n\vec{r_1} + m\vec{r_2}}{(m+n)}$$

From the above relation, we find

$$(m+n)\vec{r} - n\vec{r_1} - m\vec{r_2} = 0$$

Here we find that the sum of the coefficients of $\vec{r}$, $\vec{r_1}$, $\vec{r_2}$ is zero. If R is distinct from P and Q and at a finite distance from them, none of the coefficients in the above relation is Zero. Hence we can conclude that for three distinct collinear points R, P, Q there exists numbers l, m and n different from Zero, such that

$$l\vec{r} + m\vec{r_1} + n\vec{r_2} = 0, \; l + m + n = 0$$

Conversely, when there relations hold, the three points are Collinear.

This condition of Collinearity of three points has not been discussed in the book of NCERT.

Uniqueness of resolution of a vector in terms of its components

Suppose $\vec{OP} = \vec{r} = x\hat{i} + y\hat{j} + z\hat{k}$,

Where (x, y, z) are the components of Vector $\vec{r}$ and $\hat{i}$, $\hat{j}$ and $\hat{k}$ are unit Vectors along three mutually perpendicular directions. If possible, let $\vec{r} = x'\hat{i} + y'\hat{j} + z'\hat{k}$, Then we have

$$x\hat{i} + y\hat{j} + z\hat{k} = x'\hat{i} + y'\hat{j} + z'\hat{k}$$

$$\text{or} \quad (x - x')\hat{i} + (y - y')\hat{j} + (z - z')\hat{k} = \vec{0}$$

The right hand side is a null vector having components each equal to Zero. Again, since two Vectors are equal when their corresponding components are equal, we find from above

$$x - x' = 0, \; y - y' = 0 \text{ and } z - z' = 0$$

$$\therefore \quad x = x', \; y = y', \; z = z'$$

establishes the uniqueness of representation.

In connection with symmetrical form of equations (10.5), (10.7) to a line, one thing that requires clarification is the following. When Zero appears in one of the first two denominators, what will be the meaning of (10.5) ? To explain this consider the simultaneous equations

$$\begin{aligned} 3x + 4y - 9 &= 0 \\ x + 2y - 4 &= 0 \end{aligned} \qquad (1)$$

They yield $0.x = 2y$, $x = 3$ (2)

$$x = 3,\ y = 0$$

The equations (2) are not faulty in any way. They may be expressed as

$$\frac{x}{2} = \frac{y}{0}, \quad \frac{x}{3} = 1,$$

Or in symmetrical form as

$$\frac{x}{3} = \frac{y}{0} = \frac{1}{1} \qquad (3)$$

provided that by $\frac{y}{0}$ we do not mean $y \div 0$ and that (3) stands for (2). Thus $\frac{x}{3} = \frac{y}{0}$ stands for $0.x = 3.y$. When no denominator is 0, each ratio may be given the meaning in the sense of division. (3) can be obtained from (1) by cross multiplication. In solving equation by cross multiplication, or in writing equations in symmetrical form, relations of the form $\frac{a}{A} = \frac{b}{B}$ are to be interpreted as an alternative way of expressing the relation $B.a = A.b$. With this prior agreement, we write equations in symmetrical form. One very useful form of such equations is

$$\frac{x}{1} = \frac{y}{m} \quad \frac{z-c}{0}$$

In the textbook, 3-dim. geometry is treated after vector algebra. In the development of vectors, the idea of 3-dim. geometry has been used to some extent. The results arrived at in vectors by use of these concepts are being used again in 3-dim. geometry to explain the same concepts. This will no doubt, put the student in great difficulty. They will find themselves in a vicious circle. What has been written here is like this. To prove a = b in vector, we borrow the result A = B to be established later in 3-dim. geometry. Again to prove A = B subsequently in 3-dim. geometry, the argument that is given in the book is this : Since a = b in vectors, therefore A = B.

It is perhaps better to teach 3-dimensional geometry first, vectors next and then to show that geometry can be handled in a very neat and compact form by vectors. The students will then wonder in boundless pleasure.

However, the minimum change that is necessary in the order of treatment is the following: Coordinates, distance between two points, section ratio, direction cosines, angle between planes and lines in 3-dimensional geometry are to be treated first and then vectors. Equations to plane, line, circle in 3-dimensional space may be derived through vectors and are to be included in the chapter on vectors. But it is better for the students if they (i.e. plane, line, etc) are given independent treatment also in 3-dimensional geometry. The formulae for distance OP or PQ in Page 471 in the textbook of NCERT have not been established in vectors. The formula $OP^2 = x^2 + y^2 + z^2$ can be established by 3-dimensional geometry only. Vector can define $|\vec{OP}|$ as the positive square root of the scalar root of the scalar product $\vec{OP} \cdot \vec{OP}$. But one has to take recourse to coordinate geometry to show that this scalar represents the distance OP in the sense used in geometry. In example 9.10 -9.12, Page 421-423, the book has used the distance formula without deducing it.

The formula for d.c's (Page 473) borrows the results obtained in Page 427-428. But the second line of Page 428 is wrong because distribution law of scalar product has not been established earlier.

7. <u>Discussion of some interesting questions that may be asked by teachers to the resource persons :-</u>

1. Can you divide a Vector by another Vector ?
2. Does moment of a free Vector about a point make sense ?
3. Can we add or subtract scalar zero to a Zero Vector ?
4. What is the resolved part of Vector $\vec{b}$ in the direction of Vector $\vec{a}$? Is it a Vector or a Scalar ?
5. A line is drawn in a given direction to meet each of two skew lines. What will be the number of the points of intersection with each of the skew lines ?

References

1. Elementary Vector Analysis - C.E. Weatherburn, Orient Longman

2. Vector and Tensor Analysis - [illegible], Mcgraw Hill.

3. Vector Analysis. (Schaum Series) B. Spiegel.

LINEAR PROGRAMMING

Prepared by

Shri S. P. Das
Department of Mathematics
Bengal Engineering College, Howrah
Pin: 711103 (West Bengal)

1. Motivation of the topic

Sometimes we may face the problems like this : a business man, able to invest not more than Rs. 750.00 desires to purchase a few number of fountain pens and dot pens. A fountain pen costs Rs. 3.00 each, while a dot pen Rs. 2.00. From past experience, it is known to him that he cannot sell more than 300 pens altogether. The profits per pen in the two types of pens are Rs. .50 and Rs. .20 respectively.

Now the question is : How much of each kind of pen should he purchase? Naturally, he will try to design his purchase in such a way that within his limitations, he can get a maximum pfofit out of his investment.

It becomes a problem of maximization or minimization of some mathematical functions. Linear Programming Problem (LPP) deals with such type of problems.

The essential feature of LPP is that of linear inequality (or equality constraints and the linearity of the function to be maximized or to be minimized.

The term 'programming' means to set a plan or to design a plan in order to get a maximum or minimum functional value satisfying all the physical conditions involved in the problem.

In practice, we come across problems in which the number of constraints is not equal to the number of variables and in most of the cases. The constraining relations are in the form of inequations, so the problem ultimately reduces to solving a system of inequations.

2. Brief outline of the Content

First of all, formulate the problem mathematically, i.e. to write down a function to be maximized or to be minimized, under certain constraints. Constraints will appear either in the form of inequations or equations and along with objective function, they appear in linear form.

In next step, draw the feasible region out of the given constraints. Find the vertices or of the feasible region by solving the corresponding

equations of the sides of the feasible region. Obtain the value of the objective function at the vertices one by one. The maximum (or minimum) of these will give the maximum (or minimum) solution of the LPP. The coordinates of the corresponding vertex (or vertices) are the values of the decision variable in the extreme case.

Students or teachers may ask one question - Why are we searching only the vertices of the feasible region for its extreme value and not all points within it ? This may be explained in the following way :

Let us consider a problem of maximization. Our aim is to find a point which will correspond to a maximum value of the objective functions and at the same time satisfies all the constraints, i.e. to find a point within the feasible region which corresponds to maximum value of the objective function. For that purpose, draw the objective function for a particular value of Z. It will be a straight line. Now shift the straight line away from the origin parallel to itself by giving a series of increasing values of Z. As it is shifted away from the origin, the value of the objective function will increase. Continue this process and it will be seen that this moving straight line will just leave the feasible region by touching a single vertex or a side, i.e. two vertices. That particular point or points will be the desired solution.

For an illustration, let us take the problem :

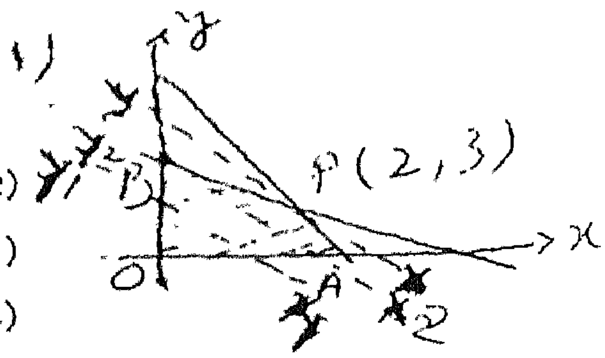
Maximize $Z = 5x + 7y$ ... (1)

Subject to

$3x + 2y \leq 12$ (2)

$2x + y \leq 13$ (3)

$x, y \geq 0$ (4)



Here OAPB is the feasible region. For fixed value of Z, objective function i parallel to $5x + 7y = 0$ (5). Plot a series of straight lines parallel to (say, X_1Y_1, X_2Y_2........ etc. All points on X_1Y_1 and within feasible regi satisfy the constraints but not correspond to a maximum value of the objectiv function because any point on X_2Y_2 gives a greater value of objective functi

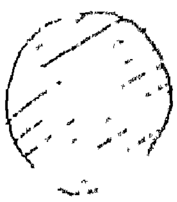
than [illegible] equation. Now let XY be one of such lines, which passes through P, the vertex farthest from the origin. On the line, only the point P (2,3) is satisfying all the constraints. Now, if it be shifted little more, none of the points on it will satisfy [illegible] constraints.

Hence [illegible] the vertices of the feasible region, is the solution of the L.P.P.

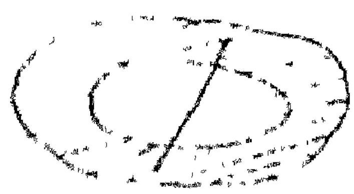
Same [illegible] applied for minimization problems also.

3. <u>Explanation of mathematical terms</u>

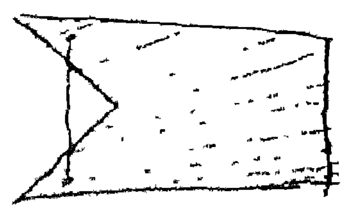
By <u>Convex set</u> we [illegible] with the property that if any two points of the set [illegible] line, the line segment lies fully on the set. For illustration [illegible] figures will help the students a lot.



A Convex set



Not a convex set



Not a convex set

Mathematically, [illegible] satisfying $x^2 + y^2 \leq 4$ is an example of convex set. While [illegible] satisfying $1 \leq x^2 + y^2 \leq 4$ is an example that this set is not convex.

<u>Feasible region</u>. The region enclosed by inequations or equations of constraints along with non-negativity of the variables is called a feasible region. Every point on or within the feasible region will satisfy the constraints.

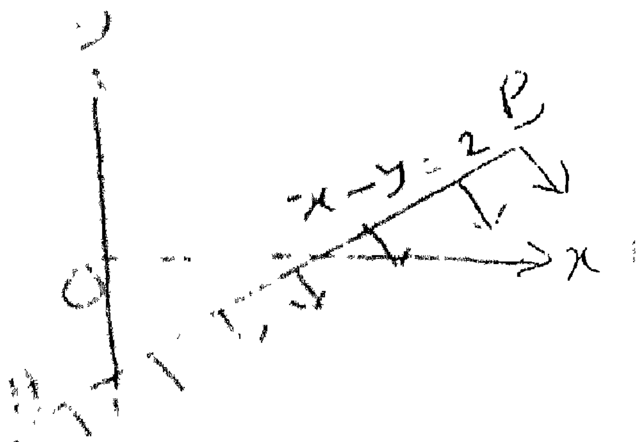
<u>Objective function</u>: is a function to be maximized or to be minimized. The variables present in the objective function are known as <u>decision variables</u>.

4. <u>Alternative approach in discussing some subtopic</u>

<u>How to draw a region corresponding to some inequations</u>:

This may be explained by with the help of an example. Let us try to draw $x - y \geq 2$.

The corresponding equation is $x - y = 2$. It can be drawn very easily. Let AB be the line. Clearly, it divides the entire xy plane into two halves.



Now in next step verify whether the origin satisfies the inequation. In this case $0 - 0 = 0 \ngtr 2$. The origin does not satisfy the inequation. Hence origin is not lying in the region. Locate in which side of AB the origin lies. Then other side of AB including the line itself will denote the required region.

If accidentally, the origin lies on AB, by taking any point P, instead of origin, we can draw the region in the same way.

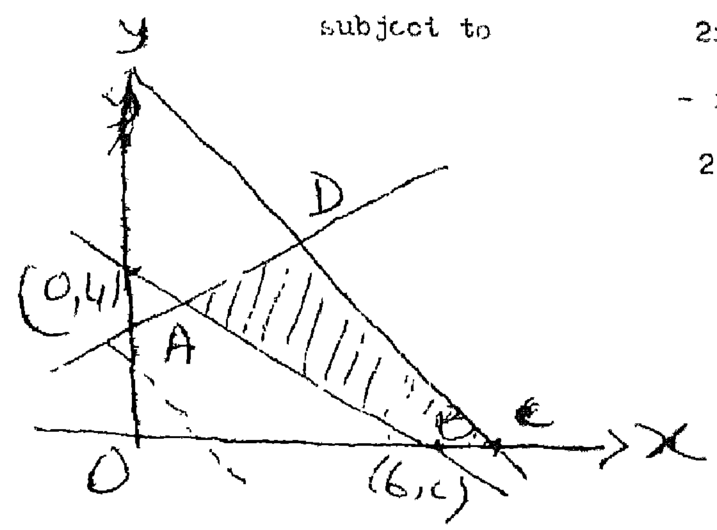
5. <u>Basic concept to be emphasised in teaching the topic</u>

In case of solving a LPP, there may be an infinite number of solutions. This will happen if the line obtained by equating the objective function to a constant, is parallel to that represented by the equation of a constraint. For example, consider the problem:

To maximize $Z = 2x + y$

subject to

$$2x + 3y \geq 12$$
$$-x + y \leq$$
$$2x + y \leq 16$$
$$x, y \geq 0$$



Here $2x + y$ = constant is parallel to $2x + y = 16$ obtained from the third constraint, i.e. parallel to CD. So the line $2x + y$ = constant when shifted away from the origin keeping parallel to itself, will leave the feasible region touching each point on the line CD.

It can be verified very easily that the objective function will give maximum value at any point on the portion of the line CD including C and D.

2. A LPP may not have a solution. This may happen in the following cases:

a) When there is no feasible region.

b) Sometimes, if the feasible region is unbounded.

3. Some problems may not have a feasible region:

$$x + y \leq 1$$
$$x \geq 2$$
$$x, y \geq 0$$

is an example in support of it. $x^2 \geq 1$ is another example of this nature.

4. Some problems may have a point region : For example

$$x + y \leq 1$$
$$x \geq 0$$
$$y \geq 1$$

is a point region.

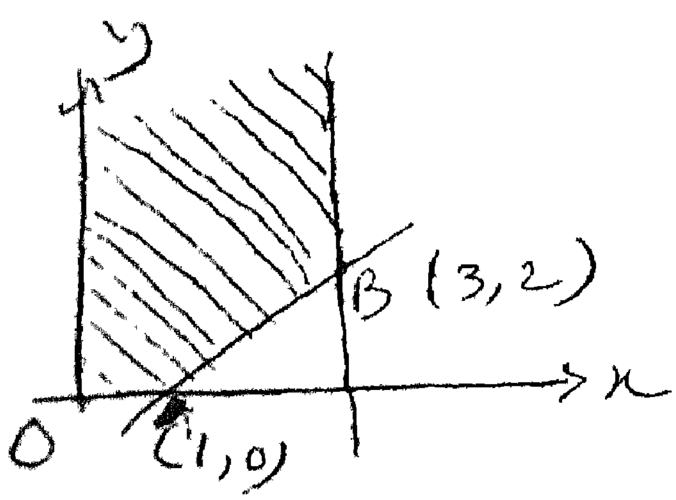
6. <u>Analysis of the possible errors that may be committed by teachers in teaching the topic</u>

In solving an optimization problem in a LPP, if the feasible region is found to be unbounded, teachers may be tempted to conclude that the solution is unbounded. But this is not always true.

This point will be clear from the following example :

Maximize $Z = 2x - y$

subject to
$$x - y \leq 1$$
$$x \leq 3$$
$$x, y \geq 0$$



As shown in the figure, the problem has an unbounded feasible region. But at the vertex B (3,2), the objective function is maximum.

It can be verified very easily that if we move along any line of feasible region towards infinity, the value of the objective function will decrease steadily as the condition of Z ... the objective function.

7. **Discussion of some enrichment material on the topic which the teachers are supposed to know**

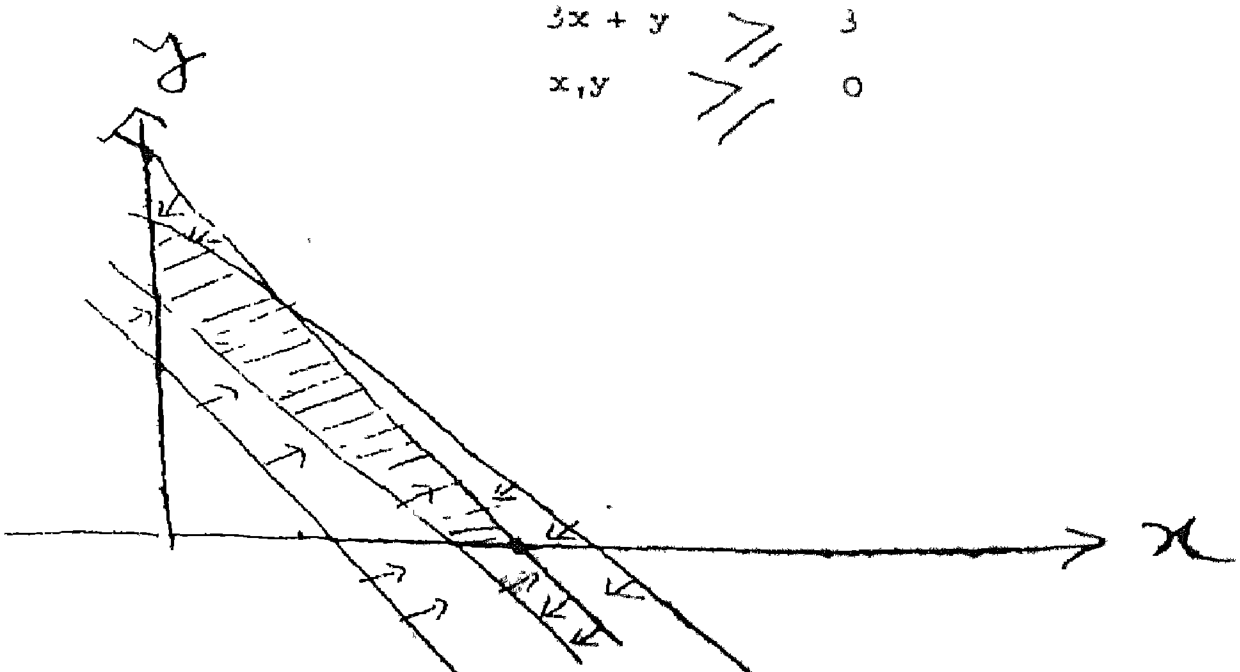
Any LPP with two variables can be solved easily by graphical method. We may also apply the same method in case of problem with 3 variables. But as the constraints are three dimensional, they will represent planes. Therefore, it will be difficult for students, particularly at this stage, to find the feasible region. Normally, a LPP with more than two variables are solved by simplex method.

In some LPP it is observed that some constraints impose no extra restriction on the feasibility of the solution and hence does not effect the solution. Such a constraint, if there be any, is called a redundant and can be neglected.

Example :

Maximize $Z = 6x + 4y$ Subject to

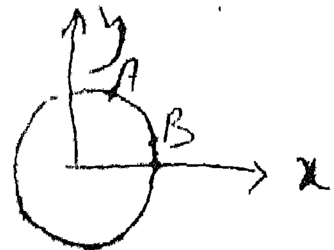
$$7x + 5y \leq 35$$
$$5x + 7y \leq 35$$
$$4x + 3y \geq 12$$
$$3x + y \geq 3$$
$$x, y \geq 0$$



The 4th constraint, i.e. $3x + y \geq 3$ has no impact to the feasible region hence it is redundant.

8. Construction of an intelligent question

Problem: Express mathematically the portion of the circumference of the circle $x^2 + y^2 = a^2$ lying between the points $A(\alpha_1, \beta_1)$ & $B(\alpha_2, \beta_2)$, $\alpha_2 > \alpha_1$, $\beta_1, \beta_2 \geq 0$.



Actually it is the portion AB of the circumference of the circle. Mathematicall

$$x^2 + y^2 = a^2, \; \alpha_1 \leq x \leq \alpha_2, \; y \geq 0$$

9. Suggested reading

1. Linear Programming ———— H. Haulee
2. Linear Programming ———— Manmohan, Gupta & Sharma
3. Operations Research ———— Taha
4. Linear Programming ———— Ghosh & Chakraborty

REGRESSION ANALYSIS

Prepared by

1. Dr. [illegible] P. [illegible]
E.C.L. (CMD's Office)
P.O [illegible]
Distt. Burdwan
West Bengal - 713333

2. Dr D. P. Sharma
Department of Mathematics
D. M. College of Science
Imphal
Manipur - 795001

1. Motivation of the subject

Suppose we want complete informations of the students of a class. The possible informations we may have of a student are:

i) Name

ii) Date of Birth

iii) Nationality

iv) Religion

v) Sex

vi) Weight

vii) Height

viii) Income status of parents

ix) Blood pressure

x) Performances in the examinations and extra-curricular activities etc.

Now if we visualise the above, we will find certain informations in terms of measurable units and the rest in the form of grading or category because a few are in quantitative form and the others in qualitative nature. Consider the following Table:

Non-measurable characteristics	Characteristics in measurable units
a) Nationality	a) Age
b) Religion	b) Weight
c) Sex status	c) Height
d) Performances in the examination	d) Parents' income status
	e) Blood pressure

Now while analysing the above tabular set of informations it is sometimes essential that elaborate statistical study of only one of the characteristics is carried on in regard to its central tendency, scatter, shape of data distribution, nature of the distribution etc. by the statistical tools Mean, Median, Mode, Range, Quartile deviation, Mean deviation, Standard deviation, Moment, Skewness, Kurtosis etc.. This type of analysis is known as Univariate analysis as only one characteristic/feature/information/variable is nurtured of the population of students of the class.

Further it may so happen that these characteristics are associated with each other by some way or other, which will lead to estimate or predict one characteristic when the other is known. When such type of mutual relationship is studied, it is called "Bivariat Analysis". Similarly when more than two characteristics are involved it is called "Multivariate Analysis".

Now, from the above set of informations the study of mutual associationship or interdependance may be tried in the following manner:

i) relationship between age and height

ii) relationship between height and weight

iii) relationship between marks scored in Mathematics and statistics in an examination

iv) relationship between age and blood pressure.

Now (i) and (ii) comes under the purview of correlation analysis; (iii) under the purview of Rank Correlation Analysis and (iv) under the purview of Regression Analysis.

Now the students may be supplied a few examples from the subjects or topics already known to them, where the variables are interrelated, e.g. established equations on Boyle's law, Charle's law, Ohm's law, Hooke's law, Newton's second law of motion, etc.

2. Brief outline of the content

a) Correlation Analysis

It is in two parts, the nature of correlation and degree of correlation. The nature of correlation can be found out drawing scatter line and observing the slope of the line with the x axis.

i) When the slope is positive the correlation is positive.

ii) When the slope is positive and the points are along a straight line the correlation is +1.

iii) When the slope is negative the correlation is negative.

iv) When the slope is negative and the points are along a line, the correlation is -1.

Again when nature and degree of correlation both are to be found out; Karl Pearson's covariance method is applied, which follows following steps:

i) Covariance (joint variation) of two variables has to be calculated (say COV (x, y))

ii) Respective standard deviation of the variables are found out (say, a_x and a_y)

iii) Then the formula is

$$r = \frac{COV\ (x,\ y)}{a_x\ a_y}$$

iv) Analysis of calculated r has to be done with proper comments, i.e. whether correlation between the variables is positive or negative; highly significant, significant, moderate or insignificant.

b) Regression Analysis

By least square method:

For a bivariate analysis of variables x and y there are two regression equations viz y on x and x on y. For y on x the steps are:

i) To consider a line $y = a + bx$ as best fit line.

ii) To solve the parameters a and b from the normal equations

$$\sum y = na + b\sum x$$
$$\sum xy = a\sum x + b\sum x^2$$

iii) Then the values of a and b are put in $y = a + bx$ to get the best fit line.

iv) Then the estimated values are calculated and the variations from the respective observed values are found out to get an idea about the fluctuations and thus fitness of the line.

Similar steps will be followed for the best fit line x on y.

c) Using formulae involving regression coefficients:

i) For the line y on x the formula is $y - \bar{y} = byx\ (x - \bar{x})$

Where $byx = \dfrac{COV\ (x,y)}{\sigma_x^2} = \dfrac{r\,a_y}{\sigma_x} = \dfrac{\dfrac{\sum xy}{n} - \dfrac{\sum x}{n}\,\dfrac{\sum y}{n}}{\sigma_x^2}$

ii) For the line x on y, the formula

$$x - \bar{x} = b_{xy}(y - \bar{y})$$

Where $b_{xy} = \frac{Cov(x, y)}{\sigma_y^2} = \frac{r\sigma_x}{\sigma_y}$

$$= \frac{\frac{\sum xy}{n} - \frac{\sum x}{n} \cdot \frac{\sum y}{n}}{\sigma_y^2}$$

3. <u>Explanation of technical/mathematical terms not properly explained in the textbook</u>

i) While stating the formula for coefficient of correlation as "Karl Pearson's Product Moment Correlation Coefficient", the term Product Moment must be explained. What is Moment ? What is Product Moment ?

Moment: It for a univariate data. If x_i ($i = 1, \ldots\ldots, n$) be the values of a variable x, then r^{th} moment of x_i about an arbitrary quantity A is defined as $\frac{\sum_{i=1}^{n}(x_i - A)^r}{n}$.

Now if A is replaced by $\bar{x}$ (the A.M of x_i) then it will be called r^{th} central moment $m_r = \frac{\sum(x_i - \bar{x})^r}{n}$,

Again if r = 1, then it is called 1st central moment of x_i and denoted as $m_1 = \frac{\sum(x_i - \bar{x})}{n}$.

Now in case of bivariate series (x_i, y_i), the r^{th} <u>product moment</u> about $\bar{x}$ and $\bar{y}$ is defined as

$$m_{rr} = \frac{\sum(x_i - \bar{x})^r(y_i - \bar{y})^r}{n}.$$

When r = 1, we get the first product moment as $m_{11} = \frac{\sum(x_i - \bar{x})(y_i - \bar{y})}{n}$

This first product moment m_{11} is called the covariance of x and y and hence the formula is also stated as Product Moment method.

ii) <u>Interpretation of 'r'</u>

The discussions in the textbook may not lead to a systematic conclusion on the interpretation of 'r'.

The following table will help to conceive as a whole:

Values of r		Interpretation on the existence of correlation between two variables
i)	$r = +1$	Perfect positive correlation
ii)	$r = -1$	Perfect negative correlation
iii)	$r = 0$	No correlation
iv)	$r > 0.36$	High degree of correlation (estimation can be relied)
v)	$r > 0.75$	Decided amount of correlation (rough estimation can be done)
vi)	$r > 0.10$	Fair degree of correlation (estimation cannot be relied)

iii) While dealing with Regression Analysis, at the outset the literal meaning of the word Regression must be known to the students. A few terms must be stressed on viz. Trend Analysis, Linear trend, Non-linear trend, Dependent variable, Independent variable.

Regression: It means stepping back to its average value. Average relationship between two variables is meant by regression.

Trend line: The line representing regression equation is known as Trend line.

Trend: It means movement of the line with respect to the axes i.e. whether the line has an upward movement or downward movement etc. When such moment is linear it is having linear trend and otherwise it shows a non-linear trend.

Dependent and Independent variables:

As there is a relation between yield of crops and manure we may say that the yield of the crops is dependent on manure i.e. a regression relation may be established as

$$\text{Yield} = f(\text{Manure})$$

Here the variable **yield** is the dependent variable which can be forecasted from the above relation. [illegible] here is independent variable.

4. <u>Alternative easier approach to any [illegible]</u>

Both in correlation and regression while tackling problems the prime and important task is to choose the proper formula and method.

a) Correlation: As for example, if only nature of correlation is to be studied only scatter diagram is sufficient.

On the other hand if both nature and degree of correlation are asked for, the proper method is Karl Pearson's coefficient of correlation formula . .. . .. Now this formula can be utilised in its short-cut form which is

$$r = \frac{\sum uv}{\sqrt{\sum u^2}\sqrt{\sum v^2}} \quad \text{where } u = x - \bar{x} \text{ and } v = y - \bar{y}$$

b) Regression: To find out the best fit line by least square method, we take help of normal equations. These normal equations can be used in simplified form by applying change of origin as per suitability of the problem. Viz. In a bivariate series where the independent variable (say x) are in A.P, the following changes may be done.

i) When n is odd,

$$X = \frac{x - \text{mid-value of } x \text{ series}}{\text{common difference of } x \text{ series}}$$

ii) When n is even,

$$X = \frac{x - (\text{A.M. of two mid-values of } x \text{ series})}{\frac{1}{2}\,(\text{common difference of } x \text{ series})}$$

Such type of change simplify the normal equations as

$$\left.\begin{aligned} \sum y &= na \\ \sum xy &= b\sum x^2 \end{aligned}\right\} \text{ for the reg. line } y = a + bx$$

and

$$\left.\begin{aligned} \sum x &= na \\ \sum xy &= b\sum y^2 \end{aligned}\right\} \text{ for the reg. line } x = a + by$$

5. Basic concept to be emphasised in teaching the topic

As it is an applied tool the students must have a clear conception about when, where and why the correlation as well as regression analysis are to be applied. Further, what are the merits and demerits of the formulae must be well explained.

a) Correlation ....

The following basic concepts must be with the students -

i) Cause and effect relationship: Although associationship is found between two variables, should we always go for correlation analysis

Examples:

1) There may be a perfect correlation found between the growth of a particular plant in a garden and the price hike of a particular commodity in the market. Now this has happened by chance. Here neither growth of the plant is cause nor price hike is the effect or vice-versa.

ii) Again another example is the relation between the variables height of living being and its age.

It may be found that there exists a correlation between the height and age till the maximum height is attained. Upto this stage cause and effect relationship is explained. But after attainment of maximum height correlation will not exist and hence cause and effect relationship no longer exists.

b) Regression

First, one must have a clear conception about the applicability of correlation and regression in different situations. If only fair degree of correlation exists between two variables, will it justify a proper linear regression equation by which estimation of dependent variable can be done against an independent variable, that is to say whether proper functional relation between two variables exists or not ?

Overall the idea must be clear that the best fit regression line will only occur when there is a perfect correlation between two variables. For an ideal best fit regression line, the value of the correlation coefficient $r = \pm 1$.

In applying formulae for correlation and regression while tackling problems, clear conception of the problem must be procured.

For Example, while applying Karl Pearson's method to determine "r" one must have an idea whether (i) the scatter line has a linear trend or not ; (ii) cause and effect relationship is found between the variables and (iii) any unduly inflated or extreme values are existing in the bivariate data affecting the value of "r".

6. Analysis of conceptual errors that may be committed by teachers in teaching the topic (including gaps/misconception in the textbook):

a) The concept of correlation and regression will be distinct if it is studied side by side as follows :

Correlation	Regression
i) It means relationship between two variables	i) It means stepping back to average value. Average relationship between two variables is stressed on.
ii) Here mutual associationship is considered, i.e. the variables are mutually dependent on each other	ii) Here the functional relation exists between two variables where one is dependent variable and the other is independent variable.
iii) Sometimes correlation exists between two variables but cause and effect relationship is not defined. So non-sense correlation exists.	iii) As it is a functional relation between two variables one dependent and other independent, cause and effect relationship is always defined.
iv) The coefficient of correlation "r" measures nature and degree of associationship.	iv) Here prediction of one variable is done when the other is known

v) The coefficient of correlation between two variables (say x and y) is symmetric, i.e. $r_{xy} = r_{yx}$. Here it does not matter which variable is dependent and which one is independent.	v) The regression coefficients are not symmetric, i.e. $b_{yx} \neq b_{xy}$ Here one variable is dependent and the other is independent.
vi) "r" is a relative measure and a pure number. It is independent of units of measurement.	vi) b_{yx} and b_{xy} are absolute measures. Its units are ~~sance~~ same as that of the respective series.
vii) Linear relationship between two variables is studied.	vii) Both linear as well as non-linear relationships are studied

b) Relation between correlation coefficient r and regression coefficients by b_{yx} and b_{xy} must be studied.

$$b_{yx} \times b_{xy} = r\frac{\sigma_y}{\sigma_x} \times r\frac{\sigma_x}{\sigma_y} = r^2$$

$$\therefore r = \pm\sqrt{b_{yx} \times b_{xy}}$$

So r is the geometric mean of the two regression coefficients. From the above, it is clear that b_{yx} and b_{xy} always have the same sign and further r, b_{yx} and b_{xy} also bear the same sign.

c) A few important property relating to r, b_{yx} and b_{xy}:

i) Two regression lines are perpendicular to each other if the product of the slopes = -1.

i.e. $b_{yx} \times \frac{1}{b_{xy}} = -1$ or, $b_{yx} = -b_{xy}$

or, $b_{yx} + b_{xy} = 0$ or, $r\frac{\sigma_y}{\sigma_x} + r\frac{\sigma_x}{\sigma_y} = 0$

or, $r\left(\frac{\sigma_y}{\sigma_x} + \frac{\sigma_x}{\sigma_y}\right) = 0$ or $r = 0$.

So $r = 0$ is the condition of perpendicularity of two regression lines.

ii) Two regression lines will be identical if the slopes are equal, i.e.

$$b_{yx} = \frac{1}{b_{xy}}$$

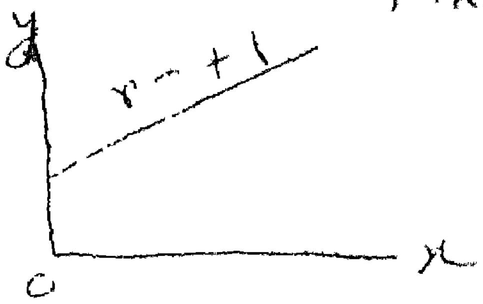
or, $b_{yx}\ b_{xy} = 1$

or $r^2 = 1$ or, $r = \pm 1$

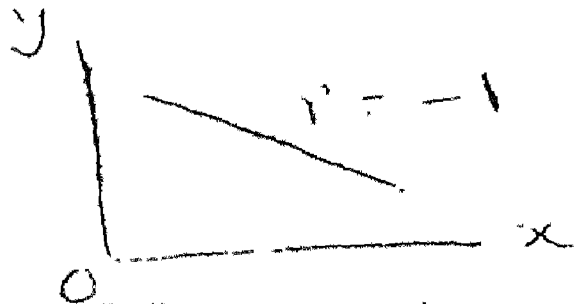
So $r = \pm 1$ is the condition of parallelism of two regression lines.

iii) The common points of two regression lines $y = a + bx$ and $x = a + by$ is the point $(\bar{x}, \bar{y})$, where $\bar{x}$ and $\bar{y}$ are the A.M. of respective series.

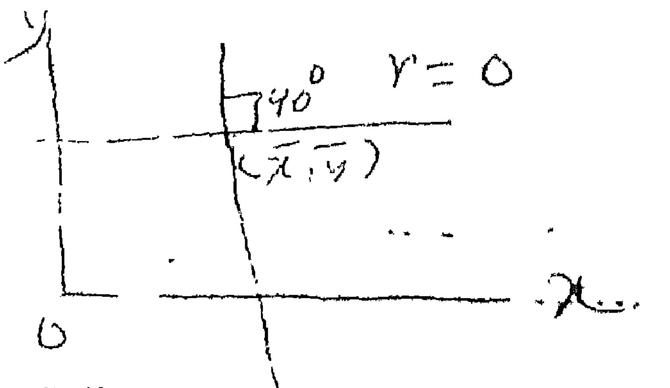
d) Geometrical representation ... regression lines: The following figures will reveal the principle that smaller is the angle between two regression lines, the greater is the degree of correlation.



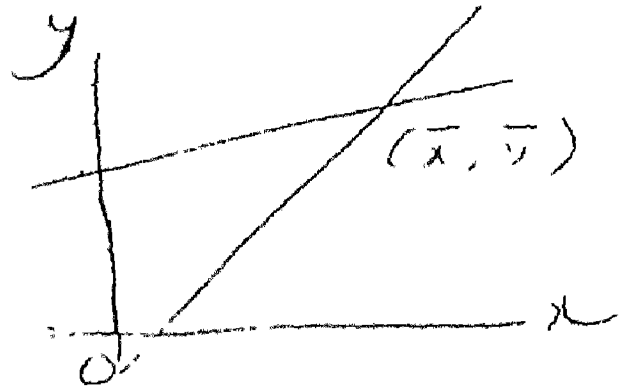
Both regression lines coincide



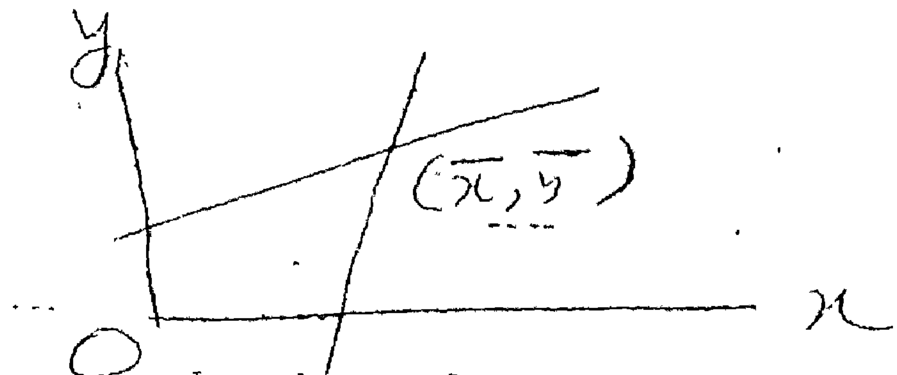
Both regression lines coincide



Both regression lines perpendicular



More degree of "r"



Less degree of "r"

e) Angle between two regression lines:

Angle between two regression lines

$$y - \bar{y} = r \frac{\sigma_y}{\sigma_x} (x - \bar{x})$$

and

$$x - \bar{x} = r \frac{\sigma_x}{\sigma_y} (y - \bar{y})$$

is given as

$$\tan\theta = \frac{m_1 - m_2}{1 - m_1 m_2}$$

$$= \frac{r\frac{\sigma_y}{\sigma_x} \sim \frac{\sigma_y}{r\sigma_x}}{1 - r\frac{\sigma_y}{\sigma_x} \times \frac{\sigma_y}{r\sigma_x}}$$

$$= \frac{\frac{\sigma_y}{\sigma_x}(1 - r^2)/r}{\frac{\sigma_x^2 + \sigma_y^2}{\sigma_x^2}} = \frac{1-r^2}{r} \cdot \frac{\sigma_y \sigma_x}{\sigma_x^2 + \sigma_y^2}$$

[Note: The gradients of the two regression lines are $m_1 = r\frac{\sigma_y}{\sigma_x}$ and $m_2 = \frac{\sigma_y}{r\sigma_x}$]

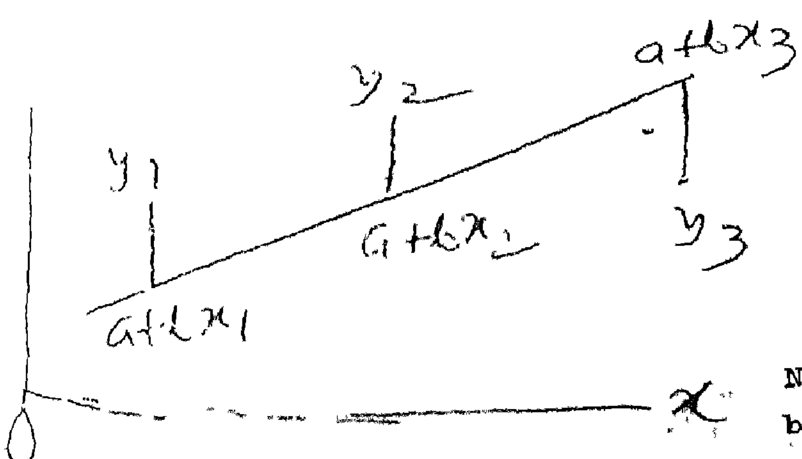
From the above the condition $r = 0$ for perpendicularity of two regression lines and $r = \pm 1$ for coincidence of two regression lines may be established.

f) Calculation of Normal Equations by Least Square Method:

For the line $y = a + bx$ to be a best fit line for the observed values (x_1, y_1), (x_2, y_2), ---, (x_n, y_n), the normal equations are calculated as follows:

Calculation of (observed y − forecasted y)

x	Observed y	forecasted y	D = (observed y − forecasted y)	S = D^2
x_1	y_1	$a + bx_1$	$y_1 - (a+bx_1)$	$(y_1-a-bx_1)^2$
x_n	y_n	$a + bx_n$	$y_n - (a+bx_n)$	$(y_n-a-bx_n)^2$
				$\sum^n S = \sum_{i=1} (y_1-a-bx_1)^2$



Vertical difference between observed value of y and estimated value of y

Now the method is the values of a and b will be such that the sum of the squares of the differences between observed and forecasted values of y become least i.e. S to be least.

Now here x, y are known quantities and so S depends on a, b i.e. S in this case will be treated as function of a and b. So when S is to be minimum

$$\frac{\partial S}{\partial a} = 0 \text{ and } \frac{\partial S}{\partial b} = 0$$ [Principle of Maxima and Minima of Diff. calculus applied]

$$\text{So } \frac{\partial S}{\partial a} = \frac{\partial}{\partial a}\left\{ \sum_{i=1}^{n} (y_i - a - bx_i)^2 \right.$$

$$= 2 \sum_{i=1}^{n} (y_i - a - bx_i) \times (-1)$$ [Partial differentiation is done]

$$= -2 \sum (y_i - a - bx_i) \quad --- (1)$$

$$\text{Similarly } \frac{\partial S}{\partial b} = -2 \sum x_i (y_i - a - bx_i) \quad ---(2)$$

From (1)

$$-2 \sum (y_i - a - bx_i) = 0$$

$$\text{Or, } \sum (y_i - a - bx_i) = 0$$

$$\text{Or, } \sum y_i - \sum a - b \sum x_i = 0$$

$$\text{Or, } \sum y_i = \sum a + b \sum x_i$$

$$\text{Or, } \sum y_i = na + b \sum x_i$$ which is the first normal equation.

From (2)

$$-2 \sum x_i (y_i - a - bx_i) = 0$$

$$\text{Or, } \sum x_i (y_i - a - bx_i) = 0$$

$$\text{Or, } \sum x_i y_i - a \sum x_i - b \sum x_i^2 = 0$$

$$\text{Or, } \sum x_i y_i = a \sum x_i + b \sum x_i^2$$

Which is the second normal equation

Similarly, we will get the normal equations of the other regression line $x = a + by$

$$\text{as } \sum x_i = na + b \sum y_i$$

$$\text{and } \sum x_i y_i = a \sum y_i + b \sum y_i^2$$

g) Rank correlation has not been mentioned in the textbook. But it is essential when comparative study between qualitative data is done e.g. comparing ... iciency tests, intelligence etc ...

or items are ranked according to their merit or efficiency and the following formulae by Spearman are applied,

a) For untied items the rank correlation coefficient

$$R = 1 - \frac{6\sum D^2}{N(N^2-1)}$$

D is the rank difference of an individual and N = number of items.

b) For tied items:

$$R = 1 - \frac{6\left\{\sum D^2 + \sum\frac{t^3-t}{12}\right\}}{N(N^2-1)},$$

Where t stands for number of individuals involved in a tie.

R satisfies the relation $-1 \leq R \leq 1$.

7. <u>Discussion of a few interesting questions that may be asked by the teachers to the Resource Persons</u>

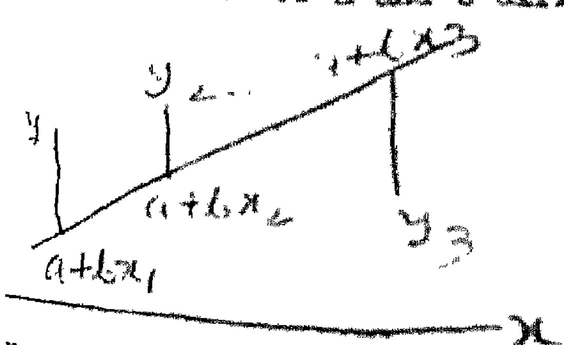
The following questions may be asked:

1) Whether two regression lines are same or not?
i.e. whether $y = a + bx$ and $x = a + by$ will help in predicting the same value.

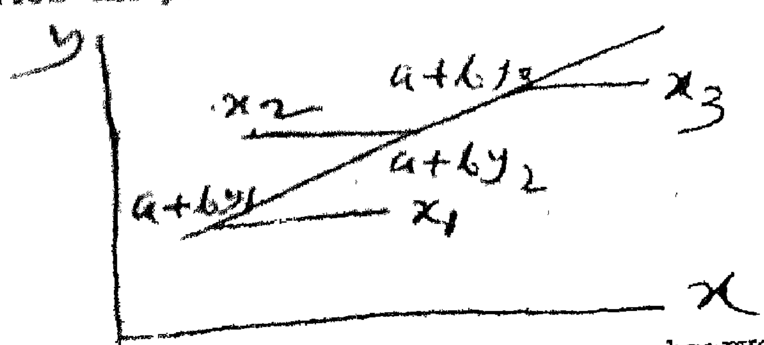
The answer is 'NO' on the following explanation:

In the regression line $y = a + bx$, we minimise the vertical difference between the observed value of y and the estimated value of y. In doing so we get the normal equation from which a and b are found out.

Now in the regression line $x = a + by$ the minimisation of differences between observed value of x and estimated value of x are done. These are horizontal displacements and the normal equations from these will give the value of a and b different from the previous a and b.



Vertical differences between observed [illegible]



Horizontal differences between observed [illegible]

ii) Another question of interest is whether any two given equations would represent the two regression lines ?

Example: Let the two equations be

$$3x + 2y = 4 \qquad --- (1)$$

and $$3y - 2x = 4 \qquad --- (2)$$

Equation (1) is expressed as y on x

i.e. $2y = 4 - 3x$

or $$y = -\frac{3}{2}x + 2 \qquad --- (3)$$

and equation (2) is expressed as x on y

i.e. $3y - 2x = 4$

or $$x = \frac{3}{2}y - 2 \qquad --- (4)$$

Now the coefficient of x in (3) is $-\frac{3}{2}$ and the coefficient of y in (2) is $\frac{3}{2}$.

Now as the two regression coefficients are of opposite signs hence we cannot consider the given two equations as regression equations.

iii) Another interesting question is given in the following example

Example: Whether $3y - 4x - 2 = 0$ and $3y - x - 6 = 0$ are regression lines?

Let the line y on x be

$$3y - 4x - 2 = 0$$

or $$y = \frac{4}{3}x + \frac{2}{3} \qquad --- (1)$$

and the line x on y be

$$3y - x - 6 = 0$$

or $$x = 3y - 6 \qquad --- (2)$$

Now both the coefficients of (1) and (2) are positive i.e. of the same sign but the product of the coefficients

$$\frac{4}{3} \times 3 = 4 > 1 \qquad ---- (3)$$

But we know that $b_{yx} \times b_{xy} = r^2$ and $-1 \leq r \leq +1$

Hence (3) states that our assumption of taking y on x and x on y are wrong.

If we consider the line

$3y - 4x - 2 = 0$ as x on y

then $x = \frac{3}{4} y - \frac{1}{2}$ --- (4)

and the line $3y - x - 6 = 0$ as y on x

then $y = \frac{1}{3} x + 2$ --- (5)

Now both the coefficients $\frac{3}{4}$ and $\frac{1}{3}$ are positive and at the same time $\frac{3}{4} \times \frac{1}{3} = \frac{1}{4} < 1$ holds good

$\therefore$ If $3y - 4x - 2 = 0$ and
$3y - x - 6 = 0$ are taken

as x on y and y on x respectively, then these will be regression lines.

8. Discussion of some enrichment materials on the topic which the teachers are supposed to know :

a) Nonlinear trend in regression analysis:

i) Method of fitting a parabolic curve :

If $y = a + bx + cx^2$ be the equation of the parabola to be fitted to the pairs of values $(x_1, y_1), (x_2, y_2), ---, (x_n, y_n)$, then the best square method may also be applied here to get three normal equations involving a, b, c as

$$\sum y = na + b\sum x + c\sum x^2$$
$$\sum xy = a\sum x + b\sum x^2 + c\sum x^3$$
$$\sum x^2 y = a\sum x^2 + b\sum x^3 + c\sum x^4$$

From the above a, b, c are calculated and put in the equation $y = a + bx + cx^2$ to get the best fit parabolic equation.

ii) Method of fitting exponential curves of the type $y = ab^x$ and $y = ae^{bx}$ as best fit.

For $y = a b^x$

or, $\log y = \log a + x \log b$ --- (1)

$\therefore$ Let $\log y = Y$

$\log a = A$

$\log b = B$

We get the equation (1) in the form

$Y = A + Bx$ which is linear in form and can be tackled as linear best fit line.

Similarly for $y = a\,e^{bx}$, the [illegible] is [illegible].

b) Teachers may have an idea about the [illegible] probable error in correlation.

Similarly the idea of standard error [illegible] best fit line.

9. <u>Construction of intelligent [illegible] concept and its solution</u>

Proper application of Regression [illegible] technique.

<u>Problem 1</u>

The following is a set of data [illegible] yield in crops [illegible] on rainfall.

	Yield in Crops (Kgs.)	[illegible] (in Cms)
Mean	508	[illegible]
Standard deviation	36	[illegible]

Here the coefficient of correlation between yield and rainfall is 0.5.

To estimate the yield of crops when rainfall is [illegible] Cms. and the rainfall when yield is 515.2 Kgs.

<u>Solution</u>: Let yield of crops is y variable and rainfall is x variable, We have

$\overline{x} = 26$, $\overline{y} = 508$,

$\sigma_x = 5$, $\sigma_y = 36$ and $r = 0.5$

The regression equation of y on x is

$y - \overline{y} = b_{yx}(x - \overline{x})$ where $b_{yx} = r\,\dfrac{\sigma_y}{\sigma_x}$

or, $y - 508 = 0.5 \times \dfrac{36}{5}\,(x - 26)$

or, $y = 3.6\,x + 414.4$

Now when x = 28 Cms

$y = 3.6 \times 28 + 414.4 = 515.2$

Now taking the regression equation of x on y i.e.

$x - \overline{x} = b_{xy}(y - \overline{y})$ when $b_{xy} = r\,\dfrac{\sigma_x}{\sigma_y}$

or, $x - 26 = 0.5 \times \dfrac{5}{36}\,(y - 508)$

or, $x = 0.069\,y - 9.052$

[illegible] = 515.2 Kgs

$x = ($ [illegible] $\times 515.2 - 9.052)$ [illegible]

$= 26.5$ Cms

So when rainfall is [illegible] Cms, the yield is forecasted as 515.2 Kgs but when yield is taken as 515.2 Kgs, then the rainfall forecasted as 26.5 Cms.

This wants justification about the applicability of the equation.

Here as yield is dependent variable and rainfall is independent, the forecasting of dependent variable is justified. The justified relation is $y = f(x)$.

Problem 2

Whether regression lines become parallel to x and y axes and can be taken as estimating lines ?

Ans: When regression lines are parallel to axes, they are perpendicular to each other.

The condition for the same is [illegible] Putting this in regression equations we get $y - \bar{y} = 0$ or, $y = \bar{y}$ —— (1)

and $x - \bar{x} = 0$ or, $x = \bar{x}$ —— (2)

So equations (1) and (2) are equations of parallel straightlines to x and y axes respectively and it is quite understandable that $y = \bar{y}$ or $x = \bar{x}$ cannot give prediction for any variable.

10. Reference Books

A) Kuebler & Smith — Statistics - A beginning

B) Leroy Folks, J — Ideas of Statistics

C) Wonnacott, R.J & Wonnacott, T.H. — Statistics discovering its power

D) -- do -- — Introductory statistics

E) Das, N.G. — Statistics

[illegible], L.B. — Teach Yourself - Statistics

— x —

NUMERICAL METHODS

Prepared by :

i) Prof. M. Mitra
Presidency College
Department of Mathematics
College Street
Calcutta - 700 073

ii) Prof. S.N. Pandey
Department of Applied Science
M.M.M. Engineering College
Gorakhpur - 273 010 (U.P.)

NUMERICAL METHODS

1. MOTIVATION

In most of the practical problems of Physics and engineering, we need a solution in numerical terms. This is achieved, in general, from the analytic solution by substituting numerical values of the data and applying the ordinary rules of arithmetic. There are number of problems where ordinary analytical methods fails to yield solutions. Also there are problems where it is simpler to obtain direct numerical solutions than to obtain the analytical solutions first and then to evaluate it for the given data. Various examples of such nature can be seen in finding the roots of transcedental equations or in solving nonlinear differential equations.

One of the simplest method is to draw a straight line or a parabola to fit a number of points in xy plane. Another is to solve the equation

$$f(x) = 0$$

by drawing the curve $y = f(x)$, and noting down the values of x at the points where y vanishes. A third is to obtain the integral of $f(x)$ by plotting the curve $y = f(x)$ on a graph paper and counting the number of squares between the curve and x-axis. These simple methods give results with low accuracy. Numerical methods can be derived from such graphical methods and they give more accurate results.

The results obtained by using numerical methods are not exact. However, calculations are performed several times to obtain a result very close to the exact value. For instance, one starts with a known

approximate solution, say, S_0. It is used to obtain a better approximation, say, S_1. Then, this solution is used to obtain a still better approximation S_2; and this process is repeated till a result of desired accuracy is reached. Therefore, the result obtained by numerical method is not exact and this method is repetitive in nature.

In this process we, therefore, deal two types of numbers - exact and approximate. For example, the numbers 1, 2, 3, ...; [illegible], $\frac{1}{3}$, $\frac{3}{2}$, ...; $\sqrt{2}$, π, e, ...; etc. written in this form are exact. But, for example, π can be given a value 3.1416. This is not exact. Also, a better value to π can be given as 3.14159265. Thus, the use of such approximations in the data give rise to errors. Therefore, the error in the result may be due to error in data or due to error in the method used or both. These errors are of various nature, and they arise due to several reasons, for example, in limiting the number of significant digits either by truncating them or by rounding them off.

Again, the rapid development of high speed digital computers and the increasing desire for numerical answers to applied problems have enhanced the demand of methods and techniques of the numerical analysis.

2. BRIEF OUTLINE OF THE CONTENT

There are several methods to solve the equations of the type $f(x) = 0$, system of linear equations and the integration of $f(x)$

numerically. However, with in the present scope of the book, it is limited to the following :

A. SOLUTION OF $f(x) = 0$.

Here $f(x)$, is assumed to be continuous and the following methods will be discussed. They are

1. The Bisection Method
2. The Method of False Position
3. Newton - Raphson Method

B. SOLUTION OF SYSTEM OF LINEAR EQUATIONS

Here the system of equations are assumed to be consistent and the number of equations and unknown variables are the same. However, the following methods are discussed taking three equations and three unknowns.

1. Gaussian Elimination Method
2. Gauss - Seidal Iteration Method

C. NUMERICAL INTEGRATION

Here again the function is assumed to be continuous, and we limit our consideration to the following two methods.

1. Trapeziodal rule.
2. Simpson's rule.

3. A BRIEF OUTLINE OF NUMERICAL METHODS

It is desirable to mention here some features of numerical methods, for instance, the significant figure, where this method is applied and so on.

Consider, for example, the following values.*

$\pi = 3.1416$

Velocity of light, $c = 299800$ km/sec

Wavelength of sodium, $D = 0.0000589$ cm.

The zeros in C and D indicate only the magnitude of the numbers and not their accuracy. The figures that remain after removing such zeros from the beginning or end are known as the significant figures. Thus π contains five significant figures while C and D contain four each (2998 and 589).

Further suppose an experiment gives the value of C = 299793.7 km/sec and the possible error in this value is ±1 km/sec. Then, it serves no purpose to write the velocity to seven figures. Only the first four figures are significant in this case. We, therefore, drop the remaining figures and round off the value to 299800.

Now we turn our attention towards the problem. A frequently occurring problem is to find the roots of the equation of the form

$$f(x) = 0 . \qquad (1)$$

If $f(x)$ is quadratic, cubic or biquadratic expression, then algebraic formulae are available to express the roots in terms of the

coefficients. But, when $f(x)$ is a polynomial of higher degree or an expression involving algebraic and transcendentals functions or so, for example,

$$1 + \cos x - 5x\ , \quad x \tan x - \cosh x\ , \quad e^{-x} - \sin x\ , \quad \text{etc.},$$

the algebraic methods are not available. Here one has recourse to approximate (numerical) methods to find the roots of such equations.

The following methods can be exployed :

1. The graphical method
2. The interaction method
3. The Bisection method
4. The method of false-position.
5. Newton-Raphson Method
6. Generalized Newton-Method
7. Muller's method
8. The quotient difference method.

SALIENT FEATURES OF THE METHODS MENTIONED IN THE CONTENT AND NOT PROPERLY EXPLAINED IN THE BOOK PAGE 619 - 653.

<u>Bisection Method</u>: Here we get successive intervals where the root lies. The length of the successive intervals is reduced to half of the length of the preceeding intervals. Here it should be remembered that in each case the root lies the interval.

Heuristically speaking one can make a rough estimate of the number of interations to be performed to obtain the desired result. For example, if we desire an accuracy of three places of decimals and choose the length of first interval to be unit, then

the length of interval after n interations should be less than .001. That is,

$$\frac{1}{2^n} < .001$$

or, $n > 10$.

So it needs to perform upto 10 interations in this case.

Further one should not use this method mechanically as evident from the book. For, consider the curve shown in the figure.

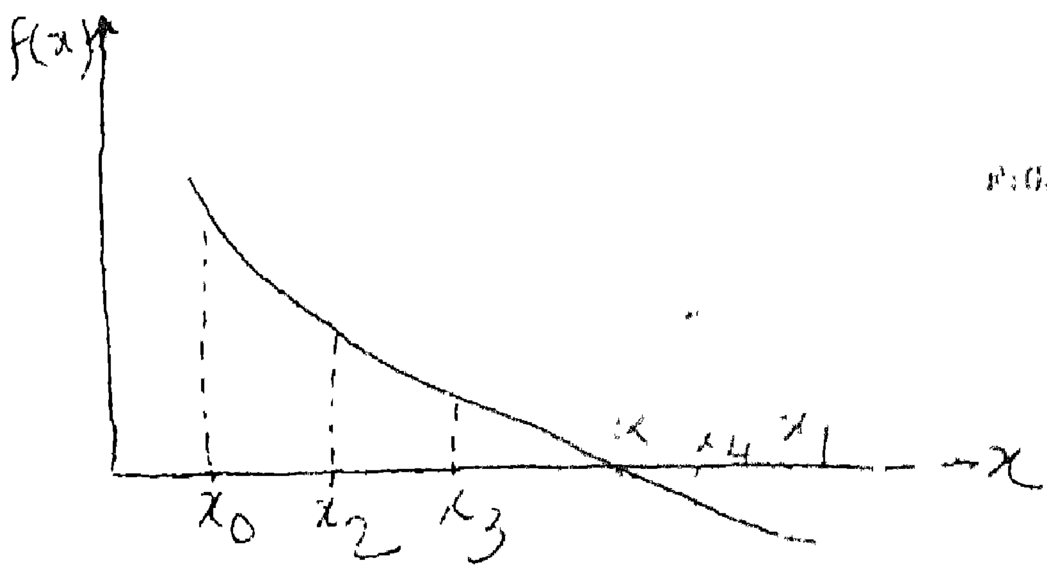


FIG. 1

The root α is close to x_1 and lies in (x_0 , x_1). The method will take several steps to reach it, because the interval is divided every time. Also, even if the root is close to $\frac{x_0 + x_1}{2}$, it requires several steps to perform to reach it.

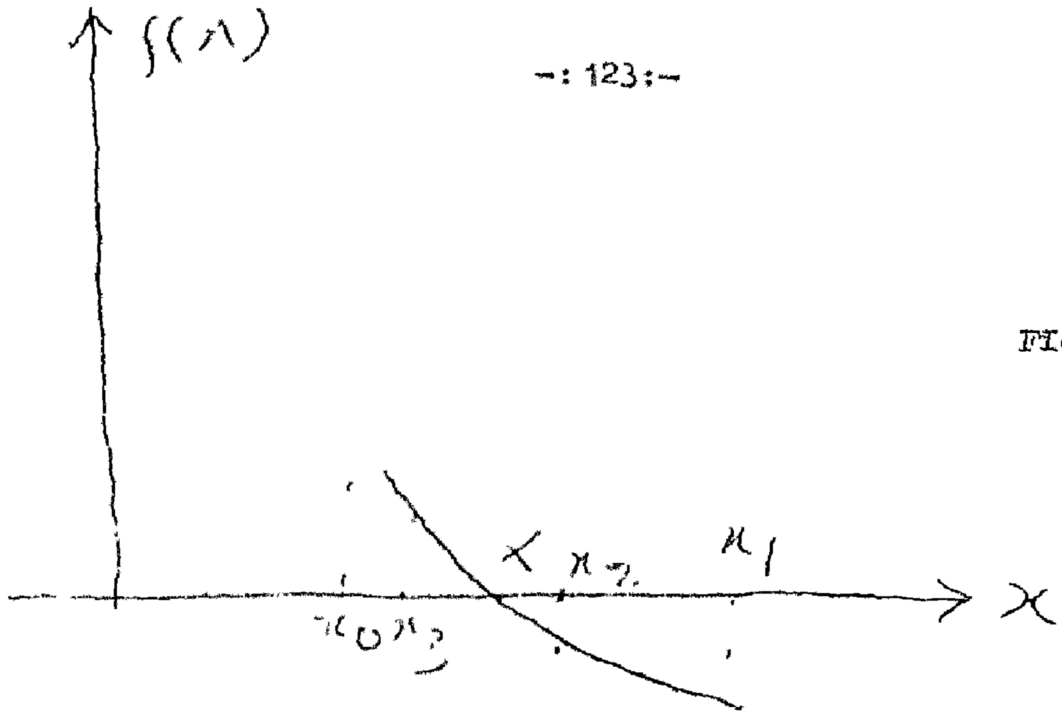


FIG. 2

This is evident from the figure 2. Therefore, it is advisable that x_0 and x_1 should be choosen close to α to its left and right respectively so that the process proceeds much faster. It is evidently clear from the example 14.1 of the book, if we follow as under :

In sixth, iteration :

$$f(x_2) = 0.00.77 \qquad \text{(This is close to } \alpha\text{)}.$$

Then, we calculate

$$f(1.795) = -0.0115$$

This means that we should choose

$$x_0 = 1.79593 \;, \quad x_1 = 1.79687$$

for the next iteration [illegible], and it gives the result of tenth iteration of the book. So, work of four iterations can be saved. Further, if we had choosen

$$f(1.796) = -.00279$$

we would have for the next iteration

$$x_0 = 1.796[illegible], \quad x_1 = 1.796$$

and this leads to a somewhat better result than the 11th iteration result of the book.

Remarks

1. The theorem (1.1) is used to get the first interval (x_0, x_1) if there is only one root, otherwise some more calculations are necessary to determine (x_0, x_1).

2. The remark 1 on p. 625 should correctly be emphasised. It is misleading. It should, for instance, be that $f(1.796) = -0.00279$ fails to satisfy the equation by -0.00279 which is a quantity close to zero.

Method of False-Position

Here the curve between two points is approximated by the line joining these points. It is better than the Bisection method as the values on the curve are used. Again, the choice of the points are important. Consider, for example, the following curves.

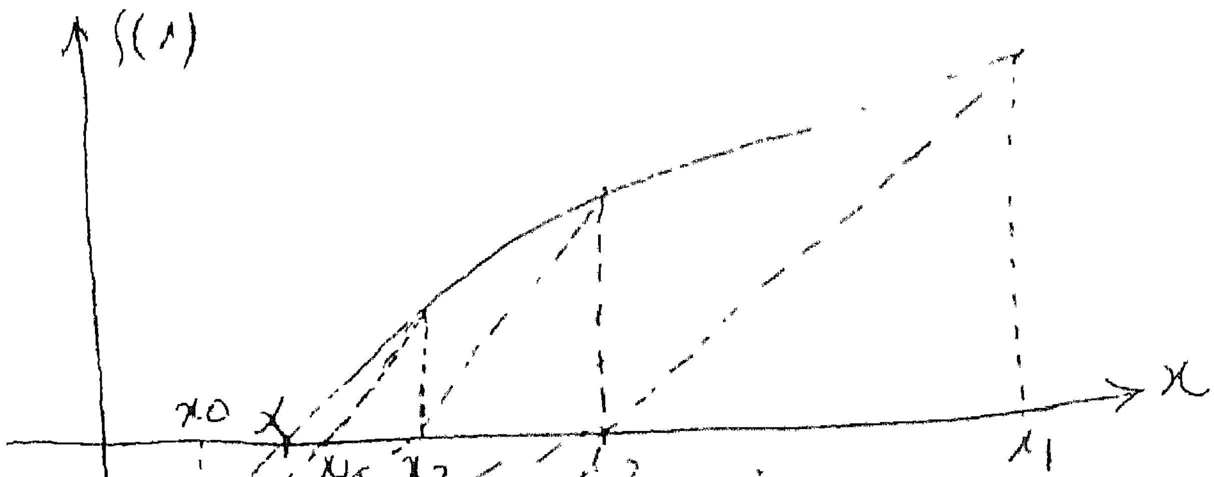


Fig 3

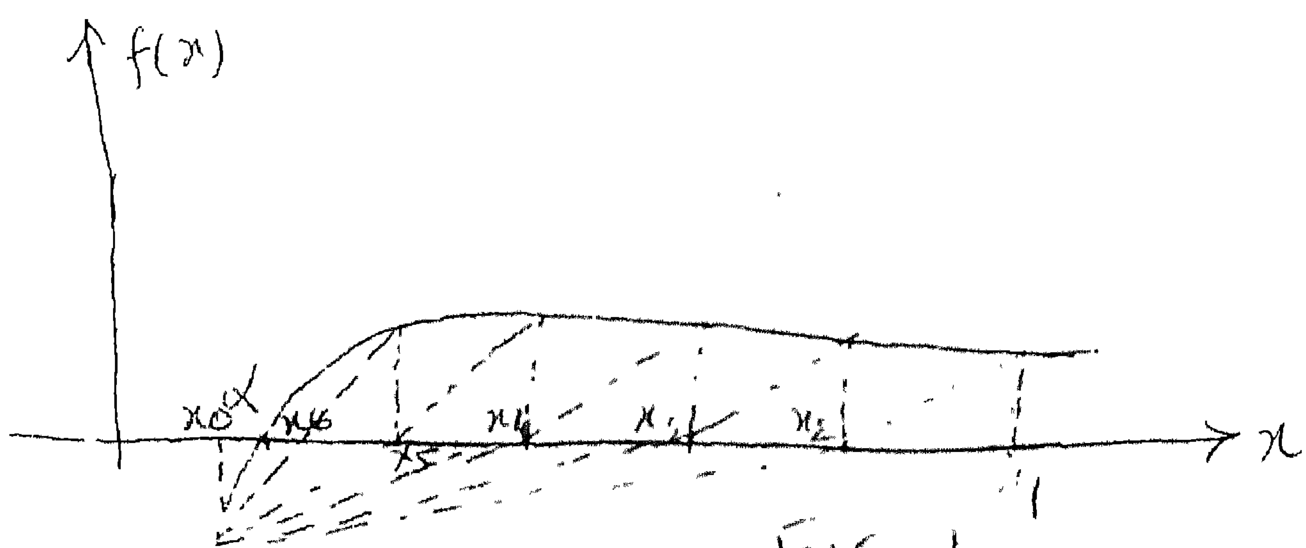


Fig 4

So, the choice of points are important, particularly when the curve is running almost parallel to the axis of x ; in this case the bisection method converges more rapidly.

NEWTON-RAPHSON METHOD

It is an improvement of method of false-position. Here again the curve is approximated by a straight line but this straight line is the tangent at one point on the curve. This method is unsuitable when the derivative at the point is small. Consider the following curves to have this point fully appreciated.

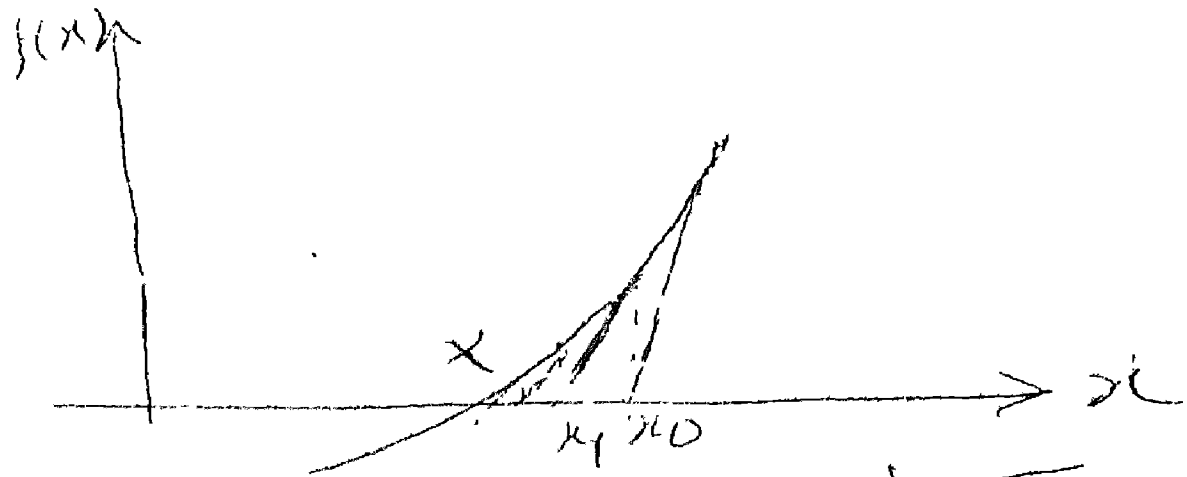


Fig 5

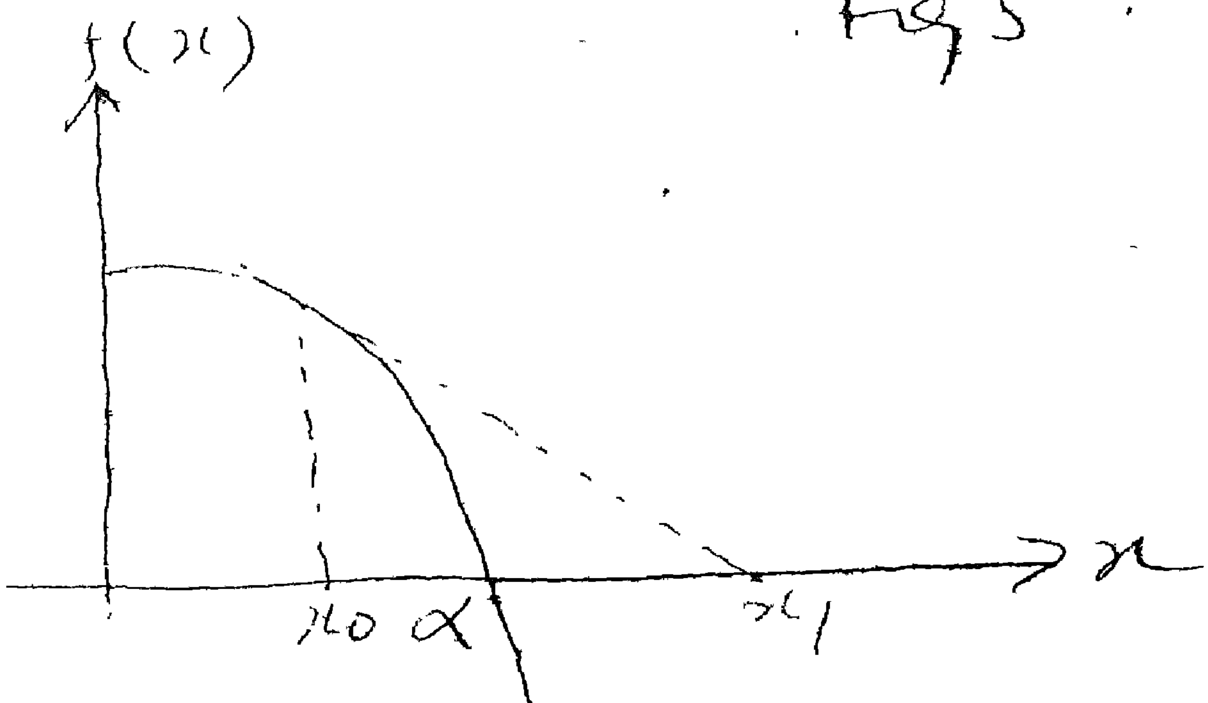


Fig 6.

GAUSS ELIMINATION METHOD

In equations (1.4.1) - (1.4.3), it should be remembered that the coefficient in the top left corner is not zero and is large compared to other coefficients of that equation, that is, (1.4.1). If it is zero or small we write the variables in different order or the equations in different order. For instance, if $a_1 = 0$ or small and b_1 is fairly large, then the equations should be written as

$$b_1 y + a_1 x + c_1 z = d_1 \quad \text{etc.}$$

This means that we should check and arrange the given equations in this fashion first before proceeding to obtain the solution. Following the method equations (1.4.1), $(1.4.2)'$, $(1.4.3)'$ will give the result on back substitution.

But, $c_3'' \neq 0$ for unique solution. If $c_3'' = 0$ and $d_3'' \neq 0$, the equation $(1.4.3)''$ is inconsistent with other two. Hence there is no solution.

Again, if $c_3'' = 0$ and $d_3'' = 0$, the equations are consistent and the values of x and y are obtained in terms of z. Z can take any arbitrary finite values. The solution is not unique.

A similar argument can be applied to the second equations, if necessary.

Remark

1. When the coefficients become too small, it is difficult to obtain the desired result. For example, equation $(1.4.3)''$ if assumes the form as

$$.00021\, z = .00012 ,$$

we get an answer correct to only two places of decimals inspite of the fact that it has been worked upto five places of decimals.

2. This method can also be used when the number of equations and variables are unequal.

GAUSS-SEIDAL METHOD

The leading diagonal terms should be fairly large. Having arranged the equations in this fashion we obtain from (14.7), (14.8), (14.9), the following :

$$\left.\begin{aligned} x &= \frac{1}{a_1}(d_1 - b_1 y - c_1 z) \\ y &= \frac{1}{b_2}(d_2 - a_2 x - c_2 z) \\ z &= \frac{1}{c_3}(d_3 - a_3 x - b_3 y). \end{aligned}\right\} \quad (2)$$

Now the iteration begins as given in the book. However, it will lead to true results if the following conditions are satisfied:

$$\left.\begin{aligned} |a_1| &> |b_1| + |c_1| \\ |b_2| &> |a_2| + |c_2| \\ |c_3| &> |a_3| + |b_3| \end{aligned}\right\} \quad \ldots\ldots (3)$$

Thus, the diagonal elements of the coefficients matrix must be fairly large compared to other elements (coefficients). If they are not, the order of the equations and variables should be changed so that the conditions (3) are satisfied.

TRAPEZOIDAL RULE

1. If the number of intervals are sufficiently large then the trapezoid by which the area in that particular interval is approximated, is getting more accurate. For example, consider

$$\int_0^1 \frac{dx}{1+x^2} = \frac{\pi}{4} = .785398 \quad \text{(6 places of decimals)} \qquad (4)$$

If (4) is evaluated with 4 equal intervals then

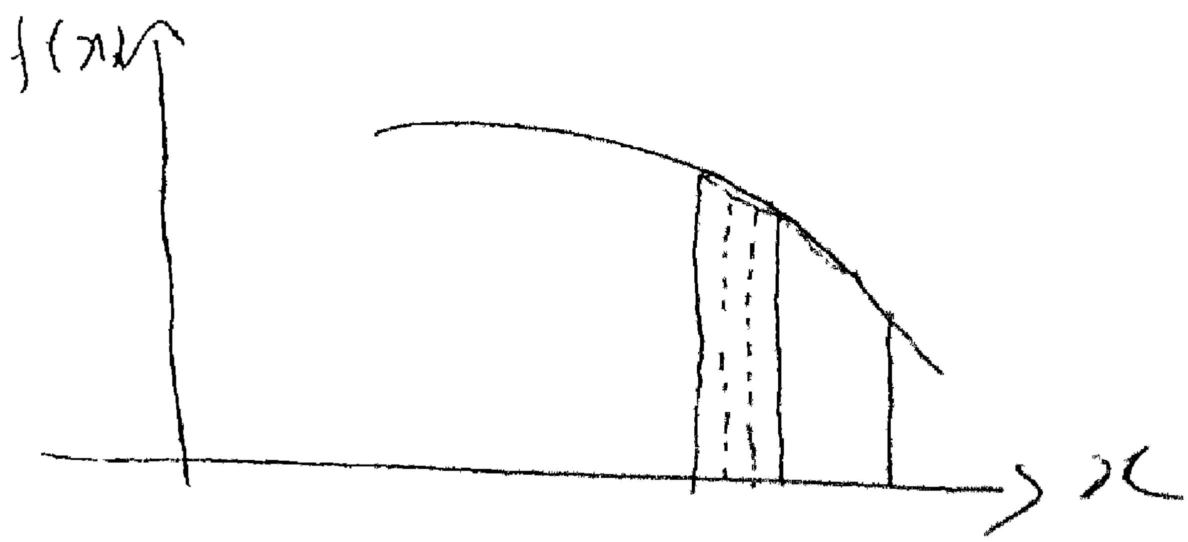
$$\int_0^1 \frac{dx}{1+x^2} = \frac{1}{8}\left[6.262\right] = .7828 \qquad (5)$$

and is evaluated with 8 equal intervals, then

$$\int_0^1 \frac{dx}{1+x^2} = \frac{1}{16}\left[12.560\right] = .78.3 \qquad (6)$$

Thus (6) is more close to $\pi/4$ than (5) :

2. The graph of the curve should approximate to straight line, in order to find the exact results by this method. For illustration, please see the following figure :



SIMPSON'S RULE

Here the curve is approximated by different parabolas. Also the result comes close to accurate value if the numbers of intervals are sufficiently more. For example, consider equation ( ), by taking 4 and 8 intervals respectively, one obtains

(i)
$$\int_0^1 \frac{dx}{1+x^2} = 0.785392 \qquad (7)$$
(4 intervals)

(ii)
$$\int_0^1 \frac{dx}{1+x^2} = 0.785398 \qquad (8)$$
(8 intervals)

A comparison can give the idea more clearly.

Remark

The interval $[a, b]$ need not be divided into equal sub-intervals. For example, consider

$$\int_{\frac{1}{4}}^{2} \frac{dx}{x} = \int_{\frac{1}{4}}^{2} \frac{dx}{x} + \int_{1}^{2} \frac{dx}{x} \qquad (9)$$

In the interval $(\frac{1}{4}, 1)$, the integrand varies rapidly whereas in $(1, 2)$ it is not varying so rapid. In such case, the interval should not be divided into equal length. Obviously, the Simpson rule will be applied to obtain the integral in two intervals $(\frac{1}{4}, 1)$ and $(1, 2)$ with different h.

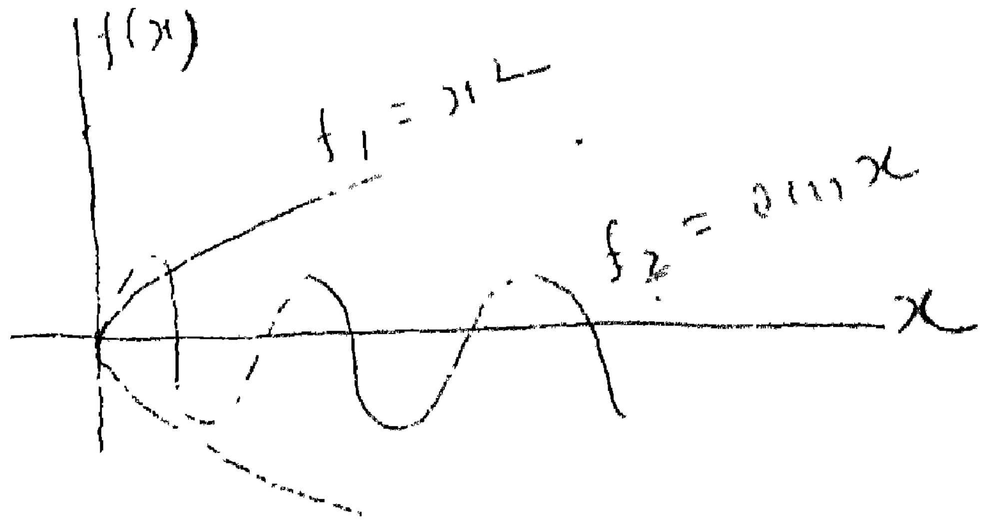
5. ALTERNATIVE AND EASIER METHODS

(a) The graphical method.

Here the functions are plotted and the point of intersection can be taken as the solution of the problem. For example, consider

$$x^2 + \sin x = 0. \qquad (10)$$

We can plot $f_1 = x^2$ and $f_2 = \sin x$ and note the point of intersection to have the approximate roots. That is,



Again, when we get the approximate value we can re-plot the curve in the vicinity of this value on large scale to have more accurate values.

(b) Iteration method.

The equation (1) can be expressed in the form

$$x = \phi(x) \qquad (11)$$

Let x_0 be an approximate value of the desired root. Then,

$$x_1 = \phi(x_0) \qquad (12)$$

Hence, the successive approximations are

$$x_2 = \phi(x_1)$$
$$- - - -$$
$$x_n = \phi(x_{n-1})$$

Here arises several questions, for example

i) Is the sequence of approximations $x_0, x_1, \dots, x_n$ always convergent, say, to limit, ξ?

ii) If it is so, will ξ be a root of the equation $x = \phi(x)$?

iii) How should we choose ϕ in order that the sequence $x_0, x_1, \dots, x_n$ converges to the root ?

It is out of the present scope to answer these questions, however, a sufficient condition for the sequence of approximations to converge is

$$|\phi'(x)| < m < 1 \qquad (12)$$

in the neighbourhood of the root.

However, the reader may look these questions by considering the equation

$$x^3 + x^2 - 1 = 0 \qquad (1.)$$

which can be put in the form (12) as

a) $x = (1 + x)^{-\frac{1}{2}}$

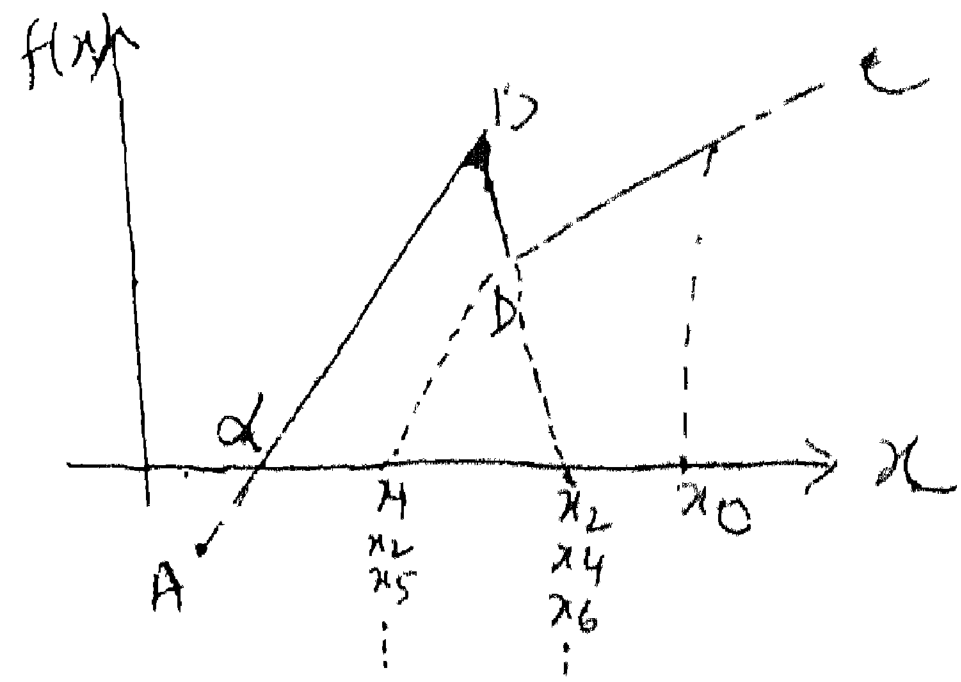
b) $x = (1 - x^3)^{\frac{1}{2}}$

c) $x = (1 - x^2)^{1/3}$

7. QUESTIONS TO TEST A PARTICULAR CONCEPT

1. Does Newton-Raphson Method always give the root if $f^1(x)$ is large in the neighbourhood of the root but $f^1(x)$ does not exist at one point only in this neighbourhood ?

The method may fail if the initial approximation x_0 is not sufficiently close to the root, α. Please see the following graph:

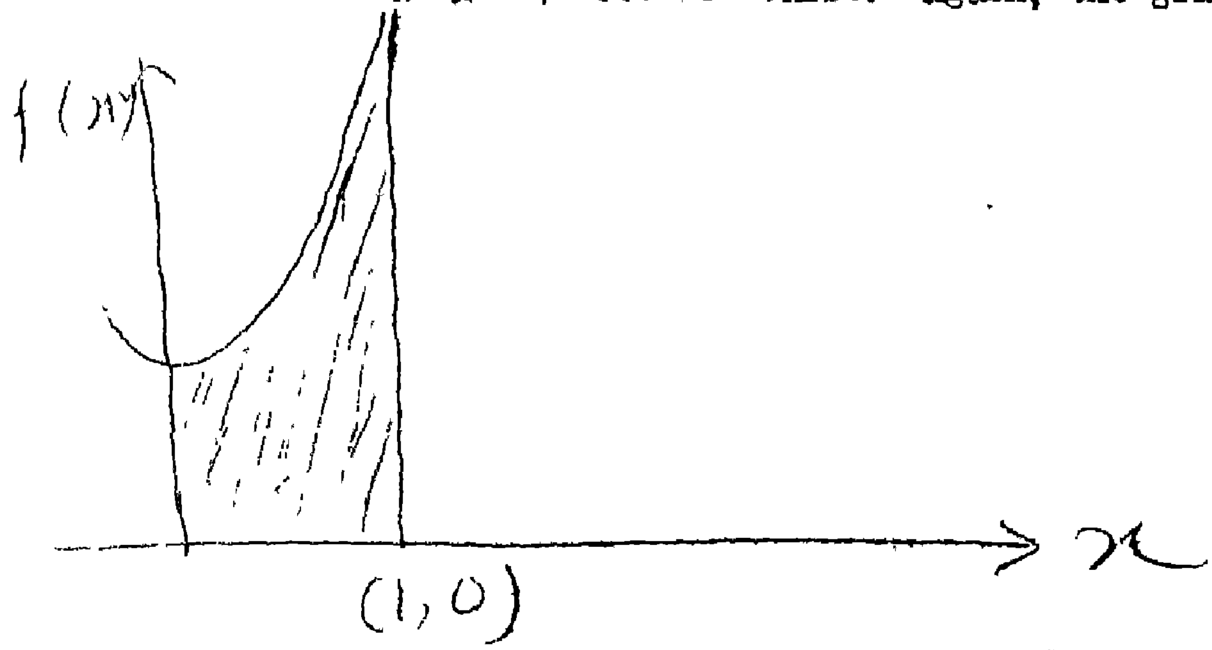


It is clear that if we start at x_0, then x_1, x_3, x_5 .... coincide as also x_2, x_4, x_6 .... Hence the process will never converge. This will happen on any point on BC. But the process will converge if the point lies on AB. Here it should be mentioned that similar examples can be constructed even if the curve is slightly curved at B and D points.

2. How can we evaluate integrals such as

$$\int_0^1 \frac{dx}{1-x^2} \quad ?$$

Here the integrand tends to infinity as x approaches unity. Trapezoidal rule and Simpson's rule are prohibited because the value of the function at x = 1 does not exist. Again, the graph is



If it is finite, we calculate

$$I_1 = \int_0^{.9} \frac{dx}{\sqrt{1-x^2}}$$

$$I_2 = \int_{.9}^{.99} \frac{dx}{\sqrt{1-x^2}}$$

$$I_3 = \int_{.99}^{.999} \frac{dx}{\sqrt{1-x^2}}$$

If we find that I_2 and I_3 become progressively smaller, we say I_1 is an approximate value of the integral ; but $I_1 + I_2$, $I_1 + I_2 + I_3$ are better approximations.

Remark

Since the power of $(1-x)$ in $\frac{1}{\sqrt{1-x^2}}$ is $\frac{1}{2}$, it can be shown from the theoretical considerations that $\int_0^1 \frac{dx}{\sqrt{1-x^2}}$ is finite.

REFERENCE (Indian Authors)

1. Chandrika Prasad — Engineering Mathematics/Mathematics for engineers. (Relevant portion)

2. S. S. Sastry — Introduction to Numerical Method.

COMPUTING

Prepared by

Dr. R. K. Bera
Department of Mathematics
A. B. N. Seal College
Coochbehar
West Bengal

COMPUTING, ALGORITHMS AND FLOW CHARTS

1. Motivation of the topic

All of us know that for one set of values of three sides of a triangle, the area of a triangle can be easily determined by some standard formula within a very short time. But if we want to calculate the areas of 100 triangles with 100 different sets of values of three sides, then the task is not very easy to do it in a very short time by the earlier methods. But now-a-days, the areas of these triangles can be computed within a time which cannot be simply conceived of a few years ago.

Another example we can cite that if we want to find the value of $F = X + Y^3$ for a set of values of (X, Y), then it can be very easily done within a very short time. But if we want to find the values of F corresponding to 1000 sets of values of (X, Y), then this cannot be done within a very short time by the earlier techniques. Once again we can say that the values of F corresponding to the 1000 sets of values of (X, Y) can be computed in an unconceivably small time, at present.

The third example, we consider, is that if we want to evaluate the integral $\int_0^1 \frac{dx}{1+x^2}$ by Simpson's One-Third rule taking only three ordinates, this can be done without much effort within a short time. But if we want to compute the same problem for 101 ordinates, then the task is not very easy and at the same time it requires much labour and time in the earlier systems. But, in the present time, this has become so easy as regards labour and time.

Hence, from the above discussions, it is clear that the computation plays a vital role in the solution of Mathematical and Scientific problems within a very short time and without giving much labour.

2. Brief outline of the content

For the computation of Mathematical, Scientific and other types of problems within a very short time and without giving much effort, the high speed computers are being used. The use of computers requires the Knowledge of Binary, octal and Hexadecimal system of numbers and their operations

tions to the decimal system.

The computer, being a machine, cannot understand our intention and language. It has its own language known as Machine language which, again, cannot be very easy to cope with for our purpose. To make the computer understand our intention and purpose we have to know some languages which computer can understand. For the sake of computer, in course of computation, we must have some knowledge of flowcharts and algorithms so that our intention and purpose may be clearly and effectively communicated to the computer to achieve our goal.

Although the computation does not require any knowledge of the computer machine, a curious reader may want to know atleast the major components of a computer.

3. Explanation of technical/mathematical terms not properly explained

The term hypothetical computer should be dropped to avoid confusion as we have already discussed the design of a computer. It is better to write a computer of 16 bits or 32 bits etc.

The difference between fixed point representation and floating point representation should be clarified. The fixed point as the name suggests has the position of the decimal point fixed. But in the case of floating point according to its name, we can change the position of the decimal point to write it in exponent form as discussed in the book.

Natural language is the language used by the human beings but, as already mentioned, the computer being a machine cannot understand, till today, the meaning of such languages as used by us. This should be explained clearly to the readers.

High level language: As the level of communication is not at per with us, a different type of language or languages will be necessary for communication to the computer machine which operates in high speed also. This languages are FORTRAN, BASIC, PASCAL, COBOL etc.

Pseudo language: The literal meaning of the word pseudo is false, that is, this language cannot be used for computer programming for communication to the computer to convey our motive and purpose. But it is helpful

for writing algorithm.

4. Alternative easier approach, if any, in discussing some subtopics

In my opinion, as our title is computation or computing, it is found that we are very much burdened with basic ideas without going much into the computation itself and the actual intention of computation is lost. It would have been better for the young mind if some examples in programming languages (not in pseudo language) are written to show how the computations are being performed. As is actually seen during computation, we do not use pseudo language.

5. Basic concept to be emphasised in teaching the topic

The concept of Binary, octal and Hexadecimal number systems is very important in relevance with switching circuits. The knowledge of Boolean Algebra is also essential. The idea of rounding off errors in numerical computation is important.

6. Analysis of conceptual errors that may be committed by teachers in teaching the topic

The relevance of reading Binary, octal and Hexadecimal system of numbers for the use of computer is not mentioned anywhere in the book.

Binary system has an application in the design of switching circuits, since a switch also has the binary characteristic of being either ON (1) or OFF (0) and since a digital computer happens to be a wiring of the switches, we can well appreciate the contribution of the binary system towards computer designing. Number is represented in the computer by a group of electrical switches which can be OFF or ON. It is, therefore, the combination of OFF (0) and ON (1) which represents a number and the fantastic computing speed of the computer is basically due to this fact. When a binary number is very large, its conversion to decimal form is time consuming and so now the provision is being made in the ultra machines to first connect binary to Octal or Hexadecimal form and then to decimal. As radix (or base) 8 is 2^3 and radix 16 is 2^4, so we first group the binary digits three at a time for conversion to Octal system and four at a time for Hexadecimal system from the

left and then we go over to the decimal system.

For example, the binary number is converted to Octal system as follows:

$$\frac{011}{3} \qquad \frac{000}{0} \qquad \frac{001}{1} \qquad \frac{111}{7} \qquad \frac{110}{6}$$

Now the Octal number 30176 can be converted to decimal number as usual.

The idea of Binary addition and subtraction as is being done within computers are not discussed anywhere in the books.

Binary addition

We can find the sum of binary numbers keeping in mind the following addition rules:

1 + 0	=	1
0 + 1	=	11
0 + 0	=	0
1 + 1	=	0 and 1 to carry.

Let us explain why we do so.

As in the decimal number system, when we come across a situation 9 + 1 which is 10, we write 0 in the corresponding place and carry 1 to the next phase (right to left). Similarly, in the binary case, we don't have a symbol 2 (which is 1 + 1), as we don't have a symbol 10 in the decimal system, so we carry 1 to the next place situated on the left of it, where spill over occurs. For example, let us add 111 to 101, we have

$$\begin{array}{r} 1\,1\,1 \\ \underline{1\,0\,1} \\ \underline{11\,0\,0} \end{array}$$

Explanation (From right to left).

1 + 1 = 0 with 1 to carry

1 + 0 = 1 + 1 (we carried) = 0 with 1 to carry

1 + 1 = 0 with 1 to carry

and 0 + 1 (we carried) = 1 and thus

$$\begin{array}{r} 1\,1\,1 \\ \underline{1\,0\,1} \\ \underline{1\,1\,0\,0} \end{array}$$

We can verify this as follows: We have $111 = 1 \times 2^2 + 1 \times 2^1 + 1 \times 2^0 = 7$ and $1\ 0\ 1 = 1 \times 2^2 + 0 \times 2^1 + 1 \times 2^0 = 5$ and $1100 = 1 \times 2^3 + 1 \times 2^2 + 0 \times 2^1 + 0 \times 2^0 = 12$, which is 7 + 5 and hence true.

Complement of numbers

In most of the computers, subtraction is carried out by adding in the minuend the complement of the number to be subtrated. This method simplifies the subtraction procedure in computers because it allows both addition and subtraction to be done by the same circuitory (the binary added). To see as how it works, let us take an example of decimal system first.

We know 135 - 125 = 10. We do the same by 9's complement method. By this method, we first find the 9's complement of the subtrahend (the number to be subtracted). The complement of a number is obtained by subtracting it from the next highest power of the radix (called 10's complementing if the radix is 10) or from the next highest power of the base less 1 (called 9's complementing if the base is 10). If the base is 2, these complements are called 2's complements and 1's complements respectively.

Thus 9's complement of 125 is $(10^4 - 1) - 125 = 999 - 125 = 874$ or in other words, we can say that if we subtract each digit of the number whose 9's complement, we are interested to find, from 9; then the outcome is the desired 9's complements. For example, 9's complements of 7465 = 2534.

Now our problem was to subtract 125 from 135.

Step I

Find the 9's complements of 125 (the subtrahend and it comes out to be 874, subtracting each digit from 9).

Step II

Add this to minuend which in this case is 135. Thus 135+874=1009.

Step III

Shift the leading digit of the sum, which is in this case 1, to the units place and find the sum; thus 1009 is now

$$\begin{array}{r} 0\ 0\ 9 \\ 1 \\ \hline 0\ 1\ 0 \end{array}$$

This is the required answer. Thus we get the same result, viz., 10 both ways.

This when applied to binary system, makes the problem highly simplified. Now, instead of finding the 9's complement as we did in the case of decimal system, we have to find 1's complement of the number which is to be subtracted and this is obtained by subtracting each digit of the number whose 1's complement we are interested to find from 1. Since in the binary system we have the digits either 1 or 0 and when we have to subtract 1 from 1, the result is 0 and when we have to subtract 0 from 1, we got 1. Thus for example, 1's complement of 100011 is 011100 which we get by simply inverting the digits, i.e. by changing 0 to 1 and 1 to 0 and this, in the computer, is done by an automatic complementor electronically.

Example: Let us for example, subtract 001101 (13) from 011100 (28)

Step I

1's complement of 001101 is 110010.

Step II

Adding the complement to the minuend

```
 011100
 110010
-------
1001110
```

Step III

Shift the leading digit of the sum obtained in step II to the unit's place and find the sum which is

```
001110
     1
------
001111
```

which is the required result which when converted to decimal is 15 and we know 28 − 13 = 15.

Major advantage of binary system is that it is applicable to most of the physical systems as they too have the binary characteristics. For example, in an electric circuit, 1 can be represented by voltage pulse and 0 by no pulse.

Another advantage of binary notation is that, with only two symbols, a limited number of addition and multiplication laws suffice and this is a big advantage over the multiplication laws in decimal system. For binary multiplication we have only to remember that $0 \times 0 = 0$, $0 \times 1 = 0$, $1 \times 0 = 0$ and $1 \times 1 = 1$.

The only <u>disadvantage</u> we have with the system is that we require a good large number of bits (Binary digits) even for the representation of a number of moderate size.

An N bit binary number is roughly equivalent to .3 N digit decimal number. For example, a 15 digit binary number 011000001111110 is required to represent 5-digit decimal number 12414 ($= .3 \times 15 = 4.5 \simeq 5$).

7. <u>Discussion of some interesting questions that may be asked by the teachers to the resource persons</u>

a) With reference to Example 13.14, a question may be asked while additing the normalised floating point numbers, we connect some numbers using its normalised form to make the exponent equal, why ?

b) With reference to Example 13.20, a question may be asked, why the final result is not round off ?

c) What is an end correction ?

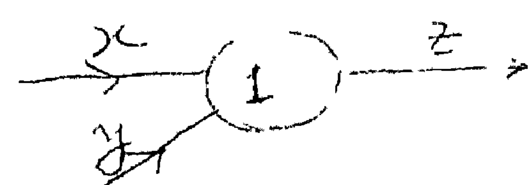
d) Is the floating point arithmetic exact ?

e) Are the operations of floating point addition and multiplication associative ?

8. <u>Discussion of some enrichment materials on the topic which the teachers are supposed to know</u>

In Brookon Algebra logical statements are converted to numerical forms. Since logical propositions are either true or false and so if we represent truth by no '1' and falsehood by no. '0', then by this it will be possible to represent logical statements in numerical form. Boolean algebra is widely used in switching circuit analysis and design and in particular, in the design of logical circuits of digital computers. From 1 (T) and 0 (F) states, certain logic statements are evolved for the circuits used in digital computers.

In Boolean Algebra, addition sign is indicated by the OR logical function. For example, if we say that Z is true if either x is true or if y is true, this in Boolean Algebra becomes $Z = x + y$. An OR logical circuit designed as follows:



It states that a signal on the output line will appear if there is a signal on either of the input lines (1 in the circle is taken to indicate this). The algebraic expression for the operation of this OR circuit is $Z = x + y$ (i.e. $Z = X \cup Y$ in set theory). We can have more than two inputs also.

Other types of logical circuits are AND and NOT.

Example : Consider the quadratic equation:

$1.0000000\, x^2 - 4.0000000\, x + 3.9999999 = 0$. If $x_1 = x_2 = 2.0000000$ are the computed roots, then it can be verified that these are the exact roots of the equation:

$0.999999992\, x^2 - 3.999999968\, x + 3.999999968 = 0$. Since the coefficients in this equation differ from those in the former by not more than one unit in the last decimal place, we can say that the aforesaid roots are fairly good roots for the former equation.

The policy of normalizing all floating point numbers can sometimes be favourable to attempt the maximum possible accuracy obtainable for a given precision and can sometimes be potentially dangerous in that it tends to imply that the results are more accurate than they really are. For example, if the result of A - B is normalised where A = +.41 3256E + 01 and B = + .413145 E + 01, then A - B = +. 111000 E - 02. In this case, the information about the possibly greater in accuracy of the result is suppressed. This information would be retained if the result were + .000111 E + 01. In order to preserve this information, unnormalized arithmetic has been suggested as has been pointed out earlier (Ashenhurst and Metropolis, 1965) in 7.

9. Suggested Reading

a) Numerical Algorithms (E.V.Krishnamurthy & S.K.Sen)-Affiliated East-West Press Pvt Ltd.

b) Numerical Analysis -- Amritava Gupta

-- M. K. Jain.

MATHEMATICAL LOGIC AND BOOLEAN ALGEBRA

Prepared by:

Dr. M.K. Sen
North Bengal University
Department of Mathematics
Raja Ram Mohunpur
Dist. Darjeeling
West Bengal

1. <u>Motivation of the topic</u>

Symbolic logic and classical logic are two disciplines included in the curriculum of Philosophy. Why then has symbolic logic taken a least out of a text book of Philosophy and is being treated as a discipline in Mathematics ? Are symbolic logic and classical logic two different subjects ? Previously, symbolic logicians criticized classical logic as out dated and defective which philosophical logicians trained in classical logic criticised symbolic logic on the ground that it involved misconceptions about the nature of logic.

This dispute is now resolved. The logicians are now unanimous that modern symbolic logic is a development of concepts and technique which were implicit in the work of Aristotle. The three features of symbolic logic are that :

(1) it uses symbols ;

(2) it uses variables ;

(3) it uses deductive method.

Now these three characteristics of symbolic logic are also characteristics of any discipline in Mathematics. Thus the development of symbolic logic has been tied up with Mathematics and it is significant that the pioneers of the subject were either mathematicians or philosophers with a mathematical training and frame of mind.

Apart from the pioneering contribution of George Boole, an English mathematician, the other name that must be mentioned is Bertrand Russell, in 1910 in collaboration with D.N. Whitehead; he published "Principia Mathematics", a monumental work, in which symbolic logic is elaborated and made to serve as the foundation of whole of Mathematics.

While the reason for studying symbolic logic in Mathematics has been outlined, again the question arises ; has it then severed its connection with Philosophy and classical logic? The answer is no. It shares with the same traditional logic the functions of providing a method of testing the validity of arguments of ordinary language and does all the tests of classicial logics in a precise way. In this respect we can say that symbolic logic is a developed form of classical logic.

One last question. Why is switching circuit theory and Boolean algebra, apparently so diverge a field of study included in this chapter ? Historically, George Boole introduced in his book, "The laws of thought" developed an algebraic system for a systematic treatment of logic, called Boolean algebra. But surprisingly afterwards Boolean algebra has found two important applications, one is that the theory of sets along with its operations of union, interaction and complement fit in nicely as an example of a Boolean algebra. And then in the second quarter of this century, it was found that the basic properties of switching circuits along with its series and parallel networks could be adequately represented by the same algebra.

Since that time, Boolean algebra has played a significant role in the important and complicated task of designing telephone circuits and electronic computers which have entered into our daily life.

The resource person may motivate the topic by concluding with a modified remark made by Russell. England can be proud of having produced two personalities who in turn were instrumental to two revolutions in the history of mankind.

One is Isaac Newton because of whose there was the industrial revolution in the past centuries and the other one is George Boole to whose credit goes the computer revolutions that the present century is witnessing.

A resource person may again ask the teachers to find a fallacy in the following argument and say that the primary aim of logic is to test whether an argument is valid or not.

1. Hypothesis

   Deer is the fastest runner

   P.T. Usha is the fastest runner (in India).

   Conclusion

   ∴ P.T. Usha is a deer

2. Hypothesis

   The fastest runner will be awarded the President's gold medal.

   Deer is the fastest runner.

   Therefore Deer will be awarded the President's gold medal.

3. 2 Mathematical terms not explained

Symbolic logic and switch algebra are treated here as particular cases of a Boolean algebra. So the attempts should be made to make the symbols and the terms more uniform.

In switch algebra, the book uses 1 and 0 for the values of the switching variables and + and . for parallel and series connections, while in symbolic logic they use T and F for truth values of a proposition and '$\vee$' and '$\wedge$' for disjunction and conjunction.

In order to emphasize to the students that they are practically doing the same mathematics, it is desirable that the same symbol be used in both the topics. Moreover, as we shall later see, calculations become very handy and a student of mathematics will feel at home if we resort to the symbols

1, 0, + and . on both the topics.

Again in order to have a parity with the terms of Boolean algebra, and that the correct message is reached, the terms connectives, simple proposition and compound proposition should be replaced by logical constant, propositional variables and propositional functions respectively. In Boolean algebra, we use the terms constant, Boolean variable and Boolean function, and there is no reason why the students are not made familiar with the corresponding terms in symbolic logic.

In the book truth tables of compound propositions such as conjunction, disjunction etc. are given all right but it should be highlighted that these are the alternative and often convenient way of defining a function.

Lastly, the definition of Boolean algebra is given in a very casual manner. They have first cited the example of the algebra of switching circuits, defined the operations of addition and multiplication in the switch algebra and explained the different laws that they satisfy and then said that such a set is called a Boolean algebra.

But it is desired that while defining a Boolean algebra, it should be made independent of all shackles, as we do in the case of theory of groups, etc, and it should then be shown that the switch algebra, the algebra of propositions are all example of a Boolean algebra.

Alternative easier approach

For all practical purposes, there are three mathematical approaches to tackle the all-important problem of determining the validity of an argument.

(i) by truth table

(ii) by contradiction

(iii) straightaway calculating with the help of different powerful operational laws and checking if the function reduce down to 1.

If so, then it is a tautology and the argument is valid; otherwise not. The method (ii) also involves mathematical calculation but the method (iii) may initially be preferred. Below are the two examples solved mathematically.

Ex. 1

$$\begin{array}{c} p \vee q \\ \underline{\sim p} \\ q \end{array}$$

Here $f = (p \vee q) \wedge (\sim p) \longrightarrow q$

$= (p + q)\, p^1 \longrightarrow q$ (converting it into mathematical language)

$= (pp^1 + qp^1) \longrightarrow q$

$= qp^1 \longrightarrow q$ $(pp^1 = 0)$

$= (qp^1)^1 + q$ by defn.

$= q^1 + p + q$ (De Morgan)

$= p + 1$

$= 1.$

So f is a tautology and the argument is valid.

Ex. 2

$$\begin{array}{c} p \longrightarrow q \\ \underline{q} \\ p \end{array}$$

$f = ((p \rightarrow q) \wedge q) \longrightarrow p$

$= ((p \rightarrow q) \cdot q) \longrightarrow p$

$= (p^1 + q) \cdot q \longrightarrow p$

$= q \longrightarrow p$ (absorption)

$= q^1 + p$ (defn.)

$\neq 1.$

So f is not a tautology and the argument is not valid.

Ex. 3 Proof by contradiction

$$\begin{array}{l} p \vee q \\ \sim p \\ \hline \end{array}$$

This is equivalent to

$$\begin{array}{l} p \vee q \\ \sim q \\ \hline p \end{array} \qquad (p, q \rightarrow r \quad p . r' \rightarrow q')$$

$$\begin{aligned} f &= (p \vee q) \wedge (\sim q) \longrightarrow p \\ &= (p + q) . q' \longrightarrow p \\ &= pq' \longrightarrow p \\ &= (pq')' + p \\ &= p' + q + p \\ &= q + 1 = 1 \end{aligned}$$

So f is a tautology and the argument is valid.

5. Basic concepts to be emphasized

(a) Proposition

A proposition or a statement is a basic entity around which any dismension of logic revolves. 'Proposition', 'true', 'false' are undefined terms, and are primitive concepts as is true of any formal system like sets in set theory, points and lines in Euclidean geometry. By a proposition is understood a sentence which has the property that it is either true or false but can not be both; it must be free of ambiguity and grammatically it is declarative in nature. Since the requirement to be free of ambiguity is relative and varies from

person to person, one may even question the existence of a proposition. Analogy can be drawn from Euclidean geometry where a line is supposed to have no width, although no such line can be drawn on a paper with a pencil. So we will come across a knotty position, if we want to make the concept too much precise and as such it is better if the reader is asked to be a little tolerant about it and not for very fussy with it.

(b) Validity of argument

One of the main purposes of a logician is to test the validity of a conclusion of some set of premises and not the truth of it. This may apparently to a beginner seem a little surprising. It may be thought that as we debate and reason only in order to arrive at a true conclusion the main instrument of the logician should be truth rather than validity. But it should be noted that the guarantee of the truth of a conclusion requires two conditions. First, the premises from which we come to the conclusion must be true and secondly the deductions must be valid of these two, logic guarantees only the second. We are not concerned at all with the truth or falsity of a proposition which are not formally deducible from other propositions. These have to be established by means which lies outside the scope of formal logic. These two conditions are totally exclusive and there is no interconnection between these two questions. It is quite likely that we can reach a true conclusion from one or two false premises and still the argument is valid and the duty of a logician is to treat only this part of validity.

Ex. 1

Premises

1. Moon is made of diamond (F)

2. Diamond is beautiful (T)

Therefore, Moon is beautiful (T)

Ex. 2

Any integer greater than 10 is even (F)

$6 > 10$ (F)

$\therefore$ 6 is an even integer (T)

6. Conceptual error/gaps

In example 5.2, p. 193 a problem is given to state the truth values of a number of simple statements. One is :

(i) There are only finite number of rational number.

It should be clear from the outset that determination of the truth value of a simple proposition is a subject that does not fall within the purview of symbolic logic and indeed it is a matter of other disciplinary study. So we can very well take the truth value of the statement "The sun rises in the West" as (T) (1) and proceed to build a consistent logical system.

7. Interesting questions

While going through the definition of the material implication $p \to q$, one may get stuck :

Truth table of $p \rightarrow q$ is as follows :

p	q	$p \rightarrow q$
T	T	T
T	F	F
F	T	T
F	F	T

The first two rows of the above table are intuitively evident as they arise most commonly in mathematics and we use it everywhere. But why are the third and the fourth rows necessary ? A satisfactory answer to this question is that a material implication "If p, then q" is a compound proposition, that is a proposition function, and in order that the function is well defined, all the possible combination of values of the (involved) variables p and q must be taken into consideration. In other words, the function must be defined for all possible values of p and q.

Incidentally, the truth table of $p \rightarrow q$ shows that it is equivalent to the function $p' + q$. So it may be pertinent very well to define $p \rightarrow q$ as the function $p' + q$. Thus the proposition $p \rightarrow q$, does not mean, as the common sense goes, that q can be logically deduced from p. The proposition only means "not p or q" and nothing more should be read from this definition.

<u>Examples</u>:

1. p : $2 + 3 = 6$ (F)

   q : The sun rises in the east (T)

   $p \rightarrow q$ is a true proposition.

2. p : $2 + 3 = 6$ (F)

q : The sun rises in the west (F)

$p \rightarrow q$ is again a true proposition.

8. Enrichment Material

(a) Truth sets for propositions

In the Book by NCERT, there is an article 'Use of Venn diagram in logic'. In this article some examples are taken to find out the truth sets of some propositions, without going into details about it.

Truth set should be of special interest to a student of Mathematics which deal with propositions that describe properties of a given universal set U. Such propositions are called propositions over U. For example, if the universal set U is the set of all +ve integers, then " x is a multiple of 5 " and " $x^2 - 16 = 0$ " are propositions over U.

Generally, if p is a proposition over U, describing some property of the element $u \in U$, then p is either true or false, depending on the particular $u \in U$ substituted in the proposition. Thus p induces on U a partition consisting of the subsets T_p and T_p', where T_p consists of all elements $u \in U$ for which p is true and $T_p' = U - T_p$.

That is $T_p = \{ u / u \in U ; p \text{ is true} \}$

The set T_p is called the truth set of p.

For example, the truth set of " x is a multiple of 5 " over the set of all +ve integers is

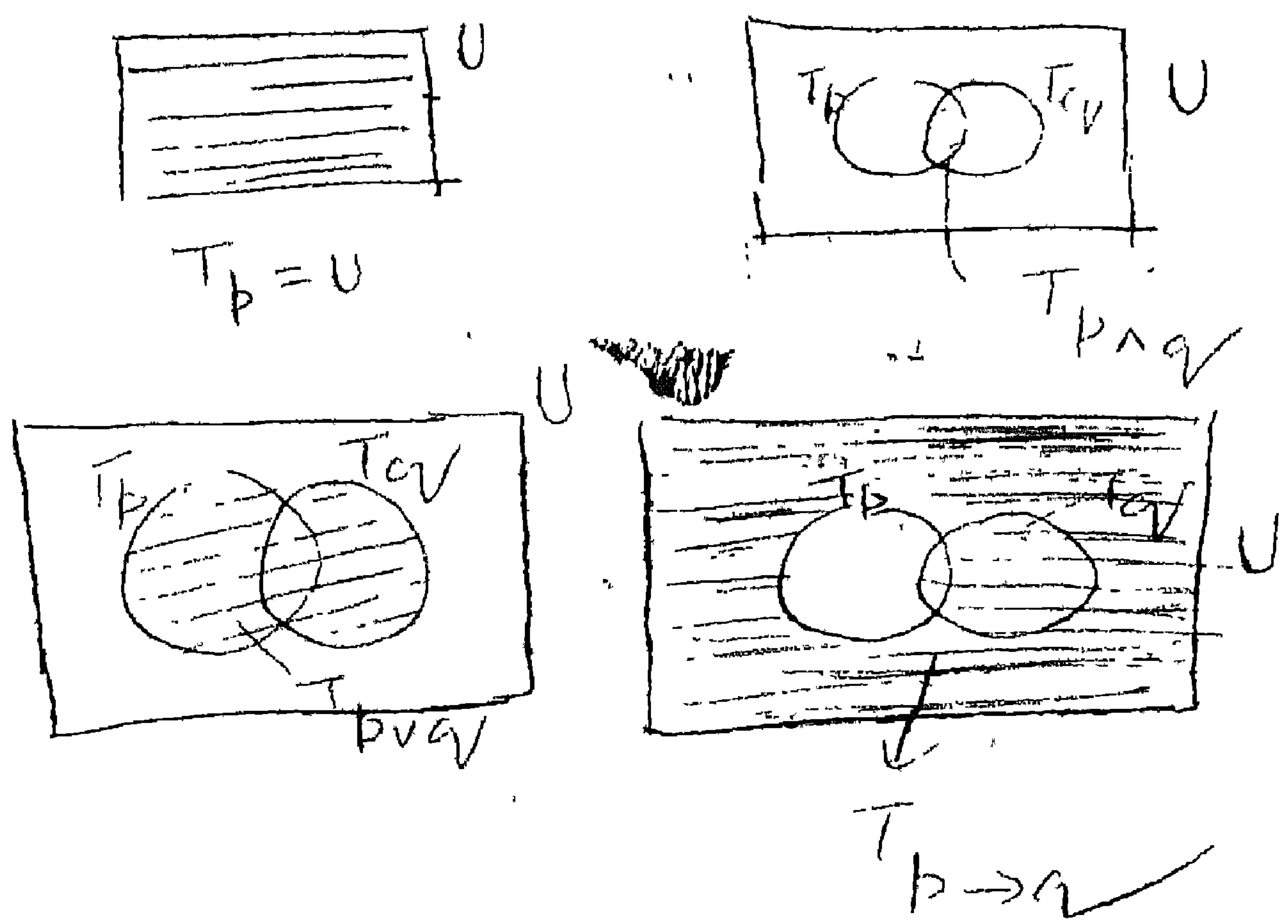
$$T_p = \{ 5, 10, 15, 20, 25, \ldots \}$$

and the truth set of " $x^2 - 16 = 0$ " over the set of all +ve integers is

$$T_p = \{ \ \}.$$

(b) The following are the truth sets for different propositions :

Tautology



Alternatively a truth set of a proposition p is a set of all logical possibilities for which p is true.

(c) Three relations in Mathematical logic

(i) $a \Longleftrightarrow b$ (a is equivalent to b)

(ii) $a \Longrightarrow b$ (a logically implies b)

(iii) a & B are inconsistent

(i.e. $a \wedge b$ is a contradiction)

In the language of sets these are :

(a) $A = B$

(b) $A \subset B$

(c) $A \cap B = \phi$

where A & B are truth sets of a and b respectively.

d. Different forms of valid arguments

1. Modus poneus

$$\begin{array}{c} p \\ p \rightarrow q \\ \hline q \end{array}$$

2. Law of syllogism

$$\begin{array}{c} p \rightarrow q \\ q \rightarrow r \\ \hline p \rightarrow r \end{array}$$

3.

$$\begin{array}{c} p \\ q \\ \hline pq \end{array}$$

4.

$$\begin{array}{c} p \cdot q \\ \hline p \end{array}$$

5.

$$\begin{array}{c} p' \\ \hline p \rightarrow q \end{array}$$

6.

$$\begin{array}{c} p + q \\ p' \\ \hline q \end{array}$$

7.

$$\begin{array}{c} p \\ \hline p + q \end{array}$$

8.

$$\begin{array}{c} q \\ \hline p \rightarrow q \end{array}$$

The valid argument of the form (7) is interesting and deserve attention.

" If Ram is watching T.V. then Ram is either watching T.V. or playing football ", is a valid argument.

REFERENCE

1. Whitesitt, Boolean algebra and its applications
2. Kemeny, Snell & Thompson, Finite Mathematics
3. R. Stoll · Set theory and logic
4. P. Suppes, Mathematical Logic.